



PROYECTO DE GRADO

Presentado ante la ilustre Universidad de Los Andes como requisito para obtener
el título de Ingeniero de Sistemas

EXTENSIÓN DEL “*BAYESIAN KNOWLEDGE TRACING*” PARA LA INCORPORACIÓN DE EMOCIONES

Por

T.S.U. Oriana Vanesa Pedroza Palomar

Tutor: Dr. Jesús Pérez

Cotutor: Dr. José Aguilar

Julio 2026

©2026 Universidad de Los Andes, Mérida, Venezuela



PROYECTO DE GRADO
CALIFICACIÓN FINAL

Título del Proyecto de Grado: "Extensión del *"BAYESIAN KNOWLEDGE TRACING"* para la incorporación de emociones"

Bachiller: Pedroza Palomar Oriana Vanesa

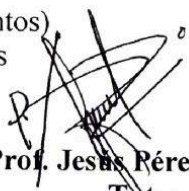
C.I: V-26.214.491

1) Calificación del (de la) Profesor(a) Tutor(a):	20ptos	(20%)
2) Calificación del manuscrito final:	20ptos	(40%)
3) Calificación de la defensa oral:	20ptos	(40%)
Calificación final:	20 ptos	

Los suscritos miembros del Jurado asignan como calificación final del Proyecto de Grado la nota de:

Observación: Los Miembros del Jurado calificador sugerimos que el presente trabajo sea accesible a la comunidad Universitaria en general, a través de la publicación de un artículo científico, donde se destaquen los principales resultados de este trabajo. En consecuencia, le otorgamos **MENCIÓN PUBLICACIÓN**.

20 (puntos)
Números


Prof. Jesús Pérez Angulo
Tutor


Prof. Junior Altamiranda Pérez
Jurado

Veinte (puntos)
Letras


Prof: Jose L. Aguilar C.
Cotutor


Prof. Wladimir Rodríguez C.
Jurado



Extensión del “*Bayesian Knowledge Tracing*” para la Incorporación de Emociones.

T.S.U. Oriana Vanesa Pedroza Palomar

Proyecto de Grado - Sistemas Computacionales

Escuela de Ingeniería de Sistemas, Universidad de Los Andes, 2026

Resumen. Los modelos de seguimiento del conocimiento, como el Bayesian Knowledge Tracing (BKT), permiten inferir el nivel de dominio de un estudiante a partir de su historial de respuestas, pero ignoran la influencia de las emociones. Las aproximaciones basadas en redes neuronales que integran información emocional para inferir el dominio requieren grandes volúmenes de datos, lo que limita su aplicabilidad en contextos educativos reales donde los datos emocionales son escasos. Este trabajo aborda dicha limitación mediante el desarrollo y evaluación de tres arquitecturas de extensión del BKT que incorporan datos emocionales. El primer enfoque incorpora el dato emocional directamente en el nodo de observación; el segundo enfoque integra el dato emocional en la predicción; y el tercer enfoque combina los dos anteriores. La experimentación se realizó sobre dos conjuntos de datos públicos ASSISTments 2012-2013 y ASSISTments Challenge 2017, incluyendo subconjuntos de 25 estudiantes con solo seis interacciones por estudiante para simular escenarios de datos limitados. Los resultados muestran que la arquitectura del segundo enfoque es la más robusta, logrando mejoras estadísticamente significativas respecto al modelo base pyBKT, en siete de los ocho escenarios evaluados, con incrementos de AUC de hasta +0,0953, y manteniéndose efectiva en escenarios de datos limitados cuando la correlación entre la emoción y rendimiento es suficientemente fuerte.

Palabras Clave: Bayesian Knowledge Tracing, seguimiento del conocimiento, emociones, concentración, modulación logística, sistemas inteligentes de tutoría.

Índice

Índice de Figuras.....	8
Índice de Tablas.....	10
Agradecimientos.....	11
Capítulo I.....	11
Introducción.....	11
1.1 Antecedentes de la Investigación.....	13
1.2 Planteamiento del problema.....	16
1.3 Justificación.....	17
1.4 Objetivos.....	18
1.4.1 Objetivo General.....	18
1.4.2 Objetivos Específicos.....	18
1.5 Metodología.....	18
1.5.1 Fase de diagnóstico.....	18
1.5.2 Fase de diseño.....	18
1.5.3 Fase de implementación.....	19
1.5.4 Fase de pruebas.....	19
1.6 Alcance.....	19
1.7 Organización del documento.....	20
Capítulo II.....	22
Marco Teórico.....	22
2.1 Emociones y rendimiento académico.....	22
2.2 Fluctuación de los estados emocionales.....	24
2.3 Modelo Bayesian Knowledge Tracing (BKT).....	26
2.3.1 Estimación de parámetros mediante el algoritmo Baum-Welch.....	30
Capítulo III.....	36
Diseño.....	36
3.1 Consideraciones.....	36
3.2 Extensiones RE-BKT.....	37
3.2.1 Premisas de diseño:.....	38
3.3 Enfoque 1: dato emocional en el nodo de observación.....	39
3.3.1 Formulación Matemática.....	39
3.3.2 Estimación de parámetros.....	46
3.4 Enfoque 2: dato emocional en la predicción.....	52
3.4.1 Formulación Matemática.....	52

3.4.2 Estimación de parámetros.....	54
3.5 Enfoque 3: Híbrido.....	55
3.5.1 Formulación Matemática.....	55
3.5.2 Estimación de parámetros.....	56
Capítulo IV.....	58
Experimentación.....	58
4.1 Protocolo experimental.....	58
4.1.1 Definición de Escenarios de Prueba.....	61
4.1.2 Conjuntos de Datos Públicos.....	65
4.2 Identificación de perfiles emocionales.....	66
4.3 Protocolo de Evaluación.....	67
4.3.1 Métricas para cada modelo.....	68
4.4 Análisis de Resultados.....	71
4.4.1 Comparación de métricas entre modelos.....	71
4.4.1.1 Resultados sobre el Dataset 1 (ASSISTments 2012-2013).....	71
4.4.1.2 Resultados sobre el Dataset 2 (ASSISTments 2017).....	74
4.4.2 Análisis Estadístico.....	79
4.4.3 Impacto de la Optimización con Optuna.....	83
4.5 Discusión de los resultados.....	85
4.6 Comparación con trabajos previos.....	88
Capítulo V.....	91
Conclusiones.....	91
5.1 Conclusiones.....	91
5.2 Aportes.....	94
5.3 Limitaciones.....	94
5.4 Recomendaciones.....	95
5.5 Trabajo Futuro.....	96
Referencias.....	97
Apéndice.....	102
Apéndice A: Análisis Exploratorio de Datos.....	102
A.1 Preprocesamiento de los conjuntos de datos.....	102
A.1.1 Preprocesamiento del Dataset 1 (ASSISTments 2012-2013).....	102
A.1.2 Preprocesamiento del Dataset 2 (ASSISTments 2017).....	103
A.2 Análisis de Correlación.....	104
A.2.1 Dataset 1: ASSISTments 2012-2013.....	105
A.2.2 Dataset 2: ASSISTments Challenge 2017.....	107
A.3 Definición de los perfiles estudiantiles.....	108
A.3.1 Selección del algoritmo y optimización del número de clústeres.....	109

A.3.2 Metodología de agrupamiento: K-Means.....	110
A.3.3 Visualización de perfiles mediante componentes principales.....	111
Apéndice B: Evaluación de las curvas de ROC.....	116
B.1 Evaluación del clúster 0 (Dataset 1 - Global).....	116
B.2 Evaluación del clúster 1 (Dataset 1 - Global).....	117
B.3 Evaluación del clúster 0 (Dataset 1 - Controlado).....	118
B.4 Evaluación del clúster 1 (Dataset 1 - Controlado).....	119
B.5 Evaluación del clúster 0 (Dataset 2 - Global).....	120
B.6 Evaluación del clúster 1 (Dataset 2 - Global).....	121
B.7 Evaluación del clúster 0 (Dataset 2 - Controlado).....	122
B.8 Evaluación del clúster 1 (Dataset 2 - Controlado).....	123

Índice de Figuras

Fig. 2.1. Parámetros del Modelo BKT en su forma matricial.....	26
Fig. 2.2. Representación del Modelo BKT.....	28
Fig. 2.3. Relación entre los estados de conocimiento y el desempeño observable.....	28
Fig. 3.1. Representación del Modelo RE-BKT con actualización del conocimiento usando la emoción interna.....	40
Fig. 3.2. Relación entre los estados de conocimiento y la respuesta - emoción observable.....	43
Fig. 4.1. Diagrama de procesos del modelo BKT.....	58
Fig. 4.2. Diagrama de las fases de entrenamiento y predicción del modelo base.....	62
Fig. 4.3. Diagrama de las fases de entrenamiento y predicción para el Enfoque 1.....	63
Fig. 4.4. Diagrama de las fases de entrenamiento y predicción para el Enfoque 2.....	64
Fig. 4.5. Diagrama de las fases de entrenamiento y predicción para el Enfoque 3.....	65

Apéndice A

Fig. A.1. Distribución porcentual de valores nulos por variable (Dataset 1).....	103
Fig. A.2. Coeficientes de Correlación Punto-Biserial ASSISTments 2012.....	105
Fig. A.3. Gráfico de barras de correlaciones ASSISTments 2012.....	106
Fig. A.4. Coeficientes de Correlación Punto-Biserial ASSISTments 2017.....	107
Fig. A.5. Gráfico de barras de correlaciones ASSISTments 2017.....	108
Fig. A.6. Evaluación de métodos de agrupamiento sobre los datos de emociones en ASSISTments 2012.....	109
Fig. A.7. Análisis de codo (a) y silueta (b) para el agrupamiento del Dataset 1.....	110
Fig. A.8. Análisis de codo (a) y silueta (b) para el agrupamiento del Dataset 2.....	111
Fig. A.9. Proyección en el plano de PCA de los clústeres de estudiantes (k=2): (a) Agrupamiento para el subconjunto Global y (b) Agrupamiento para el subconjunto Controlado.....	112
Fig. A.10. Proyección en el plano de PCA de los clústeres de estudiantes (k=2): (a) Agrupamiento para el subconjunto Global y (b) Agrupamiento para el subconjunto Controlado.....	113

Apéndice B

Fig. B.1. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster	
---	--

0).....	116
Fig. B.2. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 1).....	117
Fig. B.3. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 0).....	118
Fig. B.4. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 1).....	119
Fig. B.5. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 0).....	120
Fig. B.6. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 1).....	121
Fig. B.7. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 0).....	122
Fig. B.8. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 1).....	123

Índice de Tablas

Tabla 3.1. Parámetros del Modelo RE-BKT con Emoción Interna en su forma matricial.....	40
Tabla 4.1. Estadística de los datasets.....	66
Tabla 4.2. Diseño de la Matriz de confusión.....	69
Tabla 4.3. Métricas de rendimiento comparativas entre el modelo base y cada arquitectura del modelo RE-BKT para los clústers del Dataset ASSISTments 2012 (Global).....	72
Tabla 4.4. Métricas de rendimiento comparativas entre el modelo base y cada arquitectura del modelo RE-BKT para los clústers del Dataset ASSISTments 2012 (Controlado).....	74
Tabla 4.5. Métricas de rendimiento comparativas entre el modelo base y cada arquitectura del modelo RE-BKT para los clústers del Dataset ASSISTments 2017 (Global).....	75
Tabla 4.6. Métricas de rendimiento comparativas entre el modelo base y cada arquitectura del modelo RE-BKT para los clústers del Dataset ASSISTments 2017 (Controlado).....	76
Tabla 4.7. Arquitectura con mejor comportamiento por escenario	77
Tabla 4.8. Resumen de la prueba de DeLong para los 4 escenarios del Dataset ASSISTments 2012 - Global.....	80
Tabla 4.9. Resumen de la prueba de DeLong para los 4 escenarios del Dataset ASSISTments 2012 - Controlado.....	81
Tabla 4.10. Resumen de la prueba de DeLong para los 4 escenarios del Dataset ASSISTments 2017 - Global.....	82
Tabla 4.11. Resumen de la prueba de DeLong para los 4 escenarios del Dataset ASSISTments 2017 - Controlado.	83
Tabla 4.12. Resultados de la optimización de hiperparámetros por arquitectura, clúster y	

conjunto de datos	84
-------------------------	----

Tabla 4.13. Comparación de modelos de seguimiento del conocimiento	89
--	----

Apéndice A

Tabla A.1. Caracterización de los perfiles estudiantiles según el rendimiento y la correlación emocional (ASSISTments 2012)	113
---	-----

Tabla A.2. Caracterización de los perfiles estudiantiles según el rendimiento y la correlación emocional (ASSISTments 2017).....	115
--	-----

Agradecimientos

Durante todos estos años había esperado exactamente este momento, donde dejaría plasmado en el tiempo mi gratitud con todos aquellos seres que me impulsaron a estar justo aquí.

Gracias a Dios, a la Virgen, a los ángeles y al universo por permitirme estar con vida, por mi salud y por mantenerme en pie en los momentos más difíciles y sombríos de esta travesía. Solamente Tú, Dios, sabes el camino que tengo escrito para recorrer y sus tiempos; en tus manos entrego toda esta formación académica y personal para poder retribuir, de una u otra manera a la sociedad.

Agradezco a mis Tíos Nubia de Palomar y Misael Palomar, los pilares de mi familia. Ustedes fueron las primeras personas que recuerdo que inculcaron en mí el valor del estudio y del amor verdadero. Gracias por sus oraciones y por creer ciegamente en mi potencial.

Agradezco a mi madre, Carmenza Palomar, eres la mujer más influyente en mi vida, de la cual he aprendido incontables cosas, te doy gracias por amarnos y enseñarnos como ser resiliente con el ejemplo cada día. A mi padre, Alexander Rivas, por su apoyo, sabiduría, amor y compañía permanente. A mi padre, Oscar Pedroza, por compartir su conocimiento, consejos, apoyo y cariño en los momentos más pertinentes.

A mi hermana Grecia Pedroza le dedicaría un sinnúmero de páginas para agradecerle su existencia indispensable en mi vida. Eres mi complemento y definitivamente sin ti no estaría escribiendo estas palabras. Gracias por ser y estar para mí en cada paso y enseñarme desde tu perspectiva inigualable la realidad de la vida.

Agradezco a mi tía Karina Palomar por ser un ejemplo de amor y academia desde la infancia. Siempre la vi cómo una persona muy inteligente a quien queríamos imitar. Eres una representación de que el esfuerzo tiene su recompensa.

A mi hermano Angel Rivas, gracias por tu presencia constante. A mi tío Carlos Palomar, gracias por tu respaldo y brindarme seguridad en cada momento. Y a todos mis demás familiares (Amanda Palomar, Alba Quintero, Maria Pedroza, Ana Sanchez) gracias por cada mensaje de ánimo, cada oración y por estar presente durante mi formación.

Quiero agradecer al primer profesor que causó en mí gran impresión de ejemplo a seguir, un profesor brillante, con una calidad humana magnífica y mi tutor, Dr. Jesús Pérez. Gracias por creer en mí cuando yo no podía hacerlo, gracias por brindarme su apoyo, enseñanza y paciencia, y por inculcarme y guiarme en el área de la Investigación.

Y de manera muy especial, quiero expandir mi gratitud al Dr. José Aguilar, quien fue el formador del Dr. Jesús Pérez y también, un tutor fundamental en mi proceso. Gracias no solo por sus oportunas observaciones durante la realización de este trabajo, sino por compartir su sabiduría con generosidad.

Agradezco a mis profesores de la academia, quienes fueron indispensables en este logro y cuyos nombres recordaré con mucho cariño. Cada uno de ustedes contribuyó de manera única en mi formación, gracias por las enseñanzas y por los consejos que iban más allá de la materia.

También debo agradecer a mi profesor y colega Ing. Carlos Izquierdo quien me ayudó en mi formación como T.S.U y quien sigue apoyándome ahora como profesional. Su ayuda fue clave para poder continuar estudiando y lograr esta meta universitaria, gracias.

Les doy las gracias a todos mis compañeros de la universidad, y especialmente aquellos que fueron de mucho apoyo en distintas etapas de mi carrera: Oriana Puente, José Castro, Keinarth Aponte, Merlyn Ángulo, Carlos Puente, Jesus Valero, Ermirosa Santiago, Jessica Barrios, Daniel Monsalve, Jeremy Roa, Cherry Gota, María Guerrero, Rafael Rivas y Gustavo Vielma. Gracias porque cada uno de ustedes compartió no solo un poco de su conocimiento, sino también su tiempo y en muchos casos, su hombro cuando las noches se hacían largas y las dudas pesaban. Gracias por ayudarme a avanzar tanto en lo académico como en lo personal y espiritual, para ser una mejor persona. Muchos éxitos para ustedes.

Gracias a la Universidad de los Andes por su formación integral y a LaSDAI, donde aprendí que lo mejor que podemos hacer es compartir el conocimiento y que la crítica construye.

Capítulo I

Introducción

En los últimos años, el desarrollo de Sistemas Inteligentes de Tutoría (*Intelligent Tutoring Systems, ITS*) ha sido impulsado por los avances de la inteligencia artificial, tales como el aprendizaje automático. Estos sistemas han desempeñado un papel indispensable para lograr una educación de mayor calidad, permitiendo monitorear dinámicamente la evolución del estado de conocimiento de los estudiantes a partir de datos educativos [3].

Particularmente, el seguimiento del conocimiento (*Knowledge Tracing, KT*) es un modelo desarrollado por Corbett y Anderson [1], donde el objetivo es monitorear los estados dinámicos del conocimiento del estudiante mediante la exploración de datos relacionados con su proceso de educación. El *KT* es de gran importancia, tanto para los sistemas de aprendizaje en línea como para los estudiantes, porque permite determinar el nivel de competencia en conceptos específicos, garantizando un diagnóstico cognitivo [9]. Dentro de este ámbito, destaca el modelo tradicional de Seguimiento del Conocimiento Bayesiano (*Bayesian Knowledge Tracing, BKT*), el cual permite inferir el nivel de dominio de un estudiante sobre una habilidad específica a partir de su interacción con un sistema educativo. En relación a su estructura, este modelo está compuesto por un nodo oculto que representa estados de conocimiento latente de los estudiantes, y un nodo observable que permite estimar el conocimiento a partir del desempeño observado [12].

A su vez, Wang y sus colaboradores [3] han demostrado que la incorporación de datos emocionales (como la concentración, el aburrimiento, la confusión y la frustración) en modelos de seguimiento del conocimiento permite modelar de forma más precisa el estado del estudiante, mejorando la predicción del rendimiento académico, facilitando la creación de sistemas educativos más empáticos y adaptativos. Ellos proponen un modelo dual que permite el seguimiento tanto del estado de conocimiento como del estado emocional del estudiante, utilizando dos módulos interconectados: el “Módulo de impulso del estado del conocimiento” y el “Módulo de seguimiento del estado emocional”. No

obstante, este enfoque presenta una limitación significativa, ya que requiere una gran cantidad de datos para entrenar el modelo de manera efectiva, como los utilizados en ASSISTments 2012 [24], que incluye 29.018 estudiantes y 53.091 ejercicios, lo que puede ser un desafío en contextos educativos donde los datos de desempeño cognitivo y los datos emocionales son escasos o difíciles de recopilar.

A diferencia del modelo propuesto en [3] que requiere muchos datos, el BKT ha mostrado tener un rendimiento con pocos datos, específicamente, utilizando un tamaño de muestra de 25 estudiantes y 6 oportunidades de práctica por habilidad [4]. Por esa razón, para abordar la limitación del trabajo [3], nuestro trabajo propone el desarrollo y evaluación de tres arquitecturas de extensión del modelo clásico BKT para la incorporación de estados emocionales RE-BKT: el primer enfoque incorpora el dato emocional directamente en el nodo de observación; el segundo enfoque integra el dato emocional en la predicción; y el tercer enfoque combina ambas aproximaciones.

Los ITS están diseñados con el fin de adaptarse a las necesidades de los estudiantes para mejorar su proceso de aprendizaje. Para lograrlo, estos sistemas requieren mantener una estimación continua y precisa del conocimiento del estudiante conforme avanza en cada sesión de tutoría [11]. Por lo tanto, el seguimiento del conocimiento se convierte en una herramienta fundamental para monitorear el progreso de los estudiantes. El seguimiento del conocimiento aprovecha modelos predictivos para monitorear dinámicamente el estado de conocimiento de los estudiantes [5][6][7], siendo el BKT el modelo más popular utilizado para inferir el dominio de la habilidad a partir de los datos educativos de los estudiantes.

El BKT tradicional se modela mediante una red bayesiana dinámica, para representar la transición de sus dos estados ocultos, “*habilidad dominada*” o “*habilidad no dominada*”, donde las observaciones son las respuestas (correctas o no) del estudiante a preguntas que requieren el conocimiento de una habilidad particular [8]. En esencia, el BKT determina en qué medida un estudiante conoce una determinada habilidad a partir de su rendimiento pasado con esa habilidad, por esta razón, el modelo se compone de cuatro parámetros: probabilidad de conocer la habilidad a priori $P(L_0)$; probabilidad de que el conocimiento del estudiante sobre una habilidad pase de desconocido a conocido tras una oportunidad de aplicarla $P(T)$; probabilidad de cometer un error al aplicar una

habilidad conocida $p(S)$; y probabilidad de aplicar correctamente una habilidad desconocida $P(G)$ [2]. Aunque este modelo ha demostrado ser eficaz en entornos de aprendizaje, solo se centra en el patrón de respuestas correctas e incorrectas del estudiante, ignorando que existe información complementaria recopilada durante el proceso de aprendizaje, tales como el tiempo de respuesta o los datos afectivos, que pueden utilizarse para obtener estados de conocimiento más precisos [9].

Como se ha comentado antes, el rendimiento académico no depende exclusivamente del conocimiento cognitivo del estudiante. Las emociones también influyen, tales como la concentración, el aburrimiento, la confusión y la frustración [3]. Estas emociones inciden directamente en la forma en que el estudiante aprende y toma decisiones. A pesar de su relevancia, los métodos tradicionales de seguimiento del conocimiento no consideran explícitamente la dimensión emocional del aprendizaje. Una alternativa que puede cubrir las limitaciones más relevantes presente en estos modelos de seguimiento del conocimiento tradicionales, es una extensión del modelo clásico BKT.

1.1 Antecedentes de la Investigación

Los antecedentes de esta investigación se centran en las extensiones del modelo BKT que incorporan información más allá del rendimiento académico tradicional. Estos incluyen variables como métricas de participación estudiantil, tiempo de respuesta o afectos.

En [10], Xu y otros proponen la utilización de dispositivos económicos y portátiles de electroencefalografía (*EEG*, por sus siglas en inglés *Electroencephalogram*) para detectar estados mentales relevantes para el aprendizaje de los estudiantes, a través de ondas cerebrales. Con el propósito de alcanzar la integración de las mediciones de EEG en KT, los autores proponen que el estado de conocimiento oculto del estudiante se infiera no solo por el rendimiento observado, sino también por su estado mental medido por EEG. Por lo tanto, se agrega un nodo observable alineado temporalmente con el rendimiento del estudiante partiendo de dos enfoques principales: el primero es el método EEG-BKT que añade un nodo observable binario y modifica la tasa de aprendizaje, permitiendo que sea mayor cuando el estudiante está enfocado, sin modificar los otros parámetros. El segundo enfoque es el método EEG-LRKT que incorpora un nodo

observable multidimensional continuo, y reemplaza la tasa de aprendizaje y la tasa de olvido calculados mediante regresión logística para combinar estas medidas de EEG continuas y multidimensionales en el *KT*, lo que permite que los estados de conocimiento ocultos sean determinados por varias entradas de *EEG*. Los resultados de la investigación demostraron que ambos enfoques mejoran la predicción original del BKT, siendo EEG-LRKT significativamente superior según métricas ROC/AUC.

En [11], Schultz y otros proponen el modelo de Seguimiento de Conocimientos y Afectos (*Knowledge and Affect Tracing, KAT*) para seguir en paralelo el conocimiento y el comportamiento afectivo del estudiante. El modelo extiende el BKT al incorporar dos nodos ocultos, los cuales son: el estado de conocimiento definido como la probabilidad de dominar la habilidad en un momento dado; el estado de compromiso afectivo que permite seguir el cambio del nivel de compromiso del estudiante a lo largo del tiempo (comprometido o descomprometido). Por ende, se amplían los dos nodos observables de la siguiente manera: el nodo de la respuesta correcta ya no depende únicamente del estado de conocimiento, sino también del estado afectivo (enganchado), reemplazando los parámetros tradicionales de acierto/error por cuatro variantes condicionadas al afecto: la probabilidad de adivinar con y sin compromiso, y la probabilidad de cometer un desliz con y sin compromiso. Simultáneamente, se agrega un nodo para el comportamiento observable, el cual se clasifica en acciones específicas como “juego al sistema” o “comportamientos normales”. Esto permite al modelo distinguir si un estudiante ha “jugado” en una situación en la que estaba realmente comprometido (error afectivo), o si no estaba enganchado. En esta investigación, los autores concluyeron que el BKT tradicional fue ligeramente el mejor predictor del rendimiento futuro del estudiante en preguntas de matemáticas, pero KAT se destacó por ser el mejor en la predicción simultánea del rendimiento y los comportamientos de juego.

En la investigación llevada a cabo en [12], Wang y otros plantean una extensión que no añade directamente un nuevo nodo de Tiempo de Primera Respuesta (*First Response Time, FRT*) dentro de la estructura de red bayesiana del *KT*, sino que combina las predicciones del modelo FRT y el BKT por separado a través de un modelo de regresión lineal. En este modelo de regresión, el rendimiento del estudiante (respuesta correcta o incorrecta) se considera la variable dependiente. En cuanto a las variables

independientes, estas son las predicciones generadas por el modelo de tiempo de primera respuesta (*FRT_prediction*) y las predicciones del modelo KT (*KT_prediction*). Esta combinación, llamada por los autores “Comb”, mejora estadísticamente la precisión predictiva del BKT clásico, lo cual sugiere que respuestas rápidas pueden indicar dominio o adivinación, mientras que tiempos prolongados reflejan dificultad o conductas evasivas.

De manera similar, en [13], los autores describen la metodología de Evaluación y Valoración bayesiana (*Bayesian Evaluation and Assessment, BEA*), la cual unifica objetivos de análisis comunes en los ITS, como lo es evaluar el conocimiento del estudiante y medir el impacto de las intervenciones tutoriales. En el modelo BEA, al igual que en el modelo BKT clásico, el conocimiento del estudiante sigue siendo un nodo oculto, mientras que en la observación BEA agrega un nodo observable adicional llamado “Intervención del tutor”. Este nuevo nodo permite extender el modelo clásico del BKT para incorporar dos parámetros debido a los efectos de las intervenciones: primero, el efecto de andamiaje (*scaffold*), el cual solo sirve para ayudar al estudiante en el problema actual; segundo, efecto de enseñanza (*teach*), este modela el impacto de la intervención en el conocimiento persistente del estudiante. Como resultado, el modelo BEA determinó que las intervenciones tutoriales tienen un efecto positivo y estadísticamente significativo, ya que la probabilidad de aprendizaje es mayor cuando hay ayuda que cuando no la hay. Por consiguiente, la incorporación de información complementaria (intervención del tutor) al modelo BKT promueve el aprendizaje a largo plazo del estudiante, y proporciona un andamiaje adicional para el problema actual.

Por otro lado, Gonzalez-Brenes y otros [14] crean el método seguimiento del conocimiento de los estudiantes con reconocimiento de características (*Feature-Aware Student Knowledge Tracing, FAST*). Este método permite la integración no solo la información de acierto o error, sino que integra también múltiples características del KT, como subhabilidades (variables discretas, indicadores binarios de presencia/ausencia); dificultad de ítems y habilidad del estudiante (variables continuas), así como el historial de aciertos y errores previos, al ser capaz de representar a todos los modelos de la “Familia del seguimiento del Conocimiento”. Esto se logra al parametrizar las probabilidades de emisión y transición con regresión logística, en lugar de tablas de probabilidad condicionales. Para ello, mantienen el nodo oculto igual al del modelo

tradicional, y agregan dos nodos de características observables: como el nodo de emisión que mide cómo las características afectan la probabilidad de que el estudiante acierte o falle en un ítem dado su conocimiento y el nodo de transición que mide cómo las características afectan la probabilidad de que el estudiante aprenda u olvide entre oportunidades. En ambos nodos, las características pueden ser discretas o continuas según el caso. Los resultados indican que la predicción mejoró significativamente al BKT, alcanzando hasta un 25% en AUC en tres aplicaciones: modelado de subhabilidades mediante indicadores binarios, incorporación de dificultad de ítems y habilidad del estudiante, y uso de conocimiento experto.

Como se evidencia en las investigaciones revisadas, si bien existen avances que incorporan aspectos conductuales (como el tiempo de respuesta) y afectivos (como el compromiso o estados mentales medidos mediante EEG) en extensiones del BKT, persiste una limitación en la integración explícita de emociones como factor influyente en el modelado del conocimiento. Los datos emocionales como la concentración, el aburrimiento, la confusión y la frustración desempeñan un papel crucial en el aprendizaje, afectando directamente la atención, la retención y la motivación [3]. Partiendo de esta consideración, se proponen tres extensiones del modelo clásico BKT para incorporar estas emociones.

1.2 Planteamiento del problema

Aunque existen extensiones del BKT para incorporar información contextual, según nuestro conocimiento, no existe una extensión que considere particularmente la concentración, el aburrimiento, la confusión y la frustración, los cuales pueden alterar significativamente los procesos de adquisición y retención de conocimientos según estudios previos [3]. Según nuestro conocimiento, se han desarrollado propuestas que incorporan estados efectivos mediante el uso de redes neuronales [3][11]; sin embargo, estos presentan la limitación de que requiere volúmenes masivos de datos para su entrenamiento efectivo, resultando poco viable en contextos educativos reales donde los datos cognitivos, y sobre todo, los datos emocionales, son escasos o difíciles de recopilar. Por lo tanto, en este Proyecto de Grado se plantean tres arquitecturas de extensión del modelo clásico BKT, incorporando las emociones para mejorar las predicciones, teniendo

como propósito determinar cuál de estas configuraciones estructurales ofrece la mayor precisión predictiva en entornos de datos limitados.

1.3 Justificación

La justificación de esta investigación se fundamenta en la necesidad de mejorar los métodos tradicionales de seguimiento del conocimiento, los cuales han permitido monitorear el progreso académico de los estudiantes, pero han ignorado el papel crucial de los datos emocionales en el proceso de aprendizaje. La omisión de variables como la concentración, la frustración, la confusión y el aburrimiento deriva en diagnósticos cognitivos incompletos. La propuesta de evaluar tres enfoques diferenciados del modelo RE-BKT adquiere una relevancia metodológica particular. Ante la ausencia de un estándar arquitectónico en la literatura, este trabajo proporciona evidencia empírica sobre si las emociones deben integrarse como un componente intrínseco de la actualización del conocimiento o como un modulador logístico del rendimiento. Este análisis comparativo permite identificar la configuración óptima para maximizar la precisión predictiva sin comprometer la interpretabilidad del modelo.

A diferencia de los modelos basados en redes neuronales como en [3], que requieren grandes volúmenes de datos para evitar sobreajuste y capturar no linealidades complejas, el modelo clásico del BKT opera con solo cuatro parámetros fundamentales $P(L_0)$, $P(T)$, $P(S)$, $P(G)$. Esta estructura permite estimaciones fiables, incluso con muestras pequeñas de 25 estudiantes y 6 oportunidades de práctica por habilidad [4], por ende, dicha eficiencia se aprovechará en los experimentos con conjuntos de datos públicos, tales como ASSISTments 2012-2013 y ASSISTments Challenge 2017 [24][25].

1.4 Objetivos

1.4.1 Objetivo General

Extender el modelo “Bayesian Knowledge Tracing” para la incorporación de las emociones.

1.4.2 Objetivos Específicos

- Implementar el modelo clásico BKT utilizando conjuntos de datos públicos.
- Desarrollar una extensión del BKT considerando el dato emocional en el nodo de observación.
- Desarrollar una extensión del BKT considerando el dato emocional en la predicción.
- Desarrollar una extensión del BKT considerando el dato emocional tanto en el nodo de observación como en la predicción.
- Comparar las predicciones de los modelos propuestos con el modelo clásico BKT.

1.5 Metodología

La metodología se estructura en cuatro fases clave:

1.5.1 Fase de diagnóstico

- Revisar la literatura científica sobre extensión del modelo clásico BKT, o que vinculan la variable emocional con el rendimiento académico en los modelos tradicionales del KT.
- Revisar y evaluar la disponibilidad de conjuntos de datos públicos en educación que contengan variables relacionadas con aspectos emocionales.
- Estudiar la implementación del BKT tradicional.

1.5.2 Fase de diseño

- Diseñar el Modelo que incorpore el dato emocional en el nodo de observación.
- Diseñar el Modelo que incorpore el dato emocional en la predicción.

- Diseñar el Modelo que incorpore el dato emocional como un híbrido de los enfoques anteriores.

1.5.3 Fase de implementación

- Realizar el análisis exploratorio de datos (EDA) de los conjuntos ASSISTments 2012-2013 y ASSISTments Challenge 2017.
- Implementar el Modelo clásico BKT.
- Implementar el Modelo que incorpore el dato emocional en el nodo de observación.
- Implementar el Modelo que incorpore el dato emocional en la predicción.
- Implementar el Modelo que incorpore el dato emocional como un híbrido de los enfoques anteriores.

1.5.4 Fase de pruebas

- Evaluar el rendimiento de cada modelo extendido frente al modelo clásico BKT.

1.6 Alcance

El presente estudio se centra en el desarrollo de tres arquitecturas que extienden el modelo BKT para la incorporación de información emocional, utilizando conjuntos de datos públicos estandarizados que contienen registros cognitivos y emocionales de estudiantes. Asimismo, aunque el modelo conceptual contempla cuatro datos emocionales, la implementación se restringe a la concentración, ya que es el dato emocional que presenta mayor correlación en los datasets seleccionados. En cuanto a esto, se utilizarán inicialmente los conjuntos de datos públicos ASSISTments 2012-2013 y ASSISTments Challenge 2017 [24][25], permitiendo analizar el rendimiento académico y el comportamiento emocional de estudiantes a través de ejercicios matemáticos.

Estos conjuntos cumplen con los requisitos mínimos establecidos en la literatura revisada que sugieren que al menos 25 estudiantes y 6 oportunidades de práctica por habilidad se requieren para garantizar la robustez estadística del modelo [4]. Es pertinente

destacar que no se realizará recolección de datos primarios, sino que se trabajará exclusivamente con información secundaria disponible en estos repositorios públicos.

No obstante, es preciso señalar que el alcance del presente trabajo se limita a la fase de desarrollo y evaluación de rendimiento de cada una de las extensiones del modelo RE-BKT. Aspectos como la implementación en plataformas reales, la escalabilidad a otros dominios educativos, o la integración con sistemas de tutoría en tiempo real, quedan fuera del ámbito de este estudio y podrían considerarse como líneas futuras de investigación.

1.7 Organización del documento

Esta investigación se estructura en cinco capítulos principales:

- Capítulo I introduce la investigación, contextualizando el problema del seguimiento del conocimiento y la relevancia de incorporar las emociones. Se presentan los antecedentes, se justifica la necesidad del estudio, se definen el objetivo general y los específicos, se esboza la metodología a seguir, y se delimita el alcance del proyecto.
- Capítulo II desarrolla el marco teórico que sustenta el modelo propuesto, facilitando así la comprensión integral de la investigación. Se profundiza en la interrelación entre las emociones y el rendimiento académico, se detalla el funcionamiento del modelo Bayesian Knowledge Tracing (BKT) clásico, incluyendo sus parámetros fundamentales y el proceso de inferencia, y se explica el algoritmo *Baum-Welch* para la estimación de dichos parámetros.
- Capítulo III expone el diseño metodológico del proyecto, centrándose en el desarrollo de los modelos extendidos para incorporar el dato emocional. Se describe la formulación matemática, y la estimación de los parámetros de los tres enfoques del RE-BKT.
- Capítulo IV detalla el análisis de los datos y expone las métricas de evaluación obtenidas al comparar el rendimiento predictivo de los tres enfoques del RE-BKT frente al modelo BKT clásico sobre los conjuntos de datos seleccionados.

- Capítulo V sintetiza las conclusiones más relevantes derivadas de los resultados experimentales y discute las implicaciones teóricas de incorporar emociones al BKT. A su vez, presenta los límites, recomendaciones prácticas y posibles líneas futuras de investigación para expandir el modelo propuesto.

Capítulo II

Marco Teórico

En esta investigación se desarrollan cuatro temas centrales: primero, se explora la interrelación entre las emociones y el rendimiento académico, revisando teorías clave como la memoria dependiente del estado de ánimo, la Teoría del Control-Valor y la Teoría del Flujo, para justificar la relevancia de incorporar variables afectivas. Segundo, se explora la dinámica de los estados emocionales, analizando cómo el estudiante transita entre ellos en respuesta a eventos de aprendizaje, tal como lo describe el modelo de D'Mello y Graesser. Como tercer pilar, se detalla el funcionamiento del modelo Bayesian Knowledge Tracing (BKT) clásico, explicando su estructura, sus parámetros fundamentales y el proceso de inferencia bayesiana para actualizar el estado de conocimiento del estudiante. Finalmente, se explica el algoritmo *Baum-Welch*, método estándar para la estimación de los parámetros del BKT a partir de datos observados.

2.1 Emociones y rendimiento académico

El proceso de aprendizaje, más que una actividad puramente cognitiva, está influenciado por el estado emocional del estudiante. La psicología educativa ha reconocido que las emociones no son un subproducto de la experiencia académica, sino un componente integral que puede potenciar o dificultar la adquisición de conocimiento, la autorregulación y el rendimiento [17]. Diversas investigaciones han demostrado que las emociones influyen directamente en procesos cognitivos fundamentales como la atención, la memoria y la toma de decisiones, los cuales son pilares del rendimiento académico [3][18].

Una de las bases teóricas que sustenta la conexión entre las emociones y la cognición es el principio de la memoria dependiente del estado de ánimo (mood-state-dependent memory). Los trabajos de Bower [18] demostraron que los eventos aprendidos en un estado emocional particular se recuerdan mejor cuando la persona se encuentra nuevamente en ese mismo estado, por ende, la emoción del momento facilita el acceso a los recuerdos vinculados a ella. Este hallazgo tiene

implicaciones directas para el aprendizaje, ya que un estudiante en un estado de ánimo positivo recordará con mayor facilidad experiencias de aprendizaje placenteras, mientras que un estado de ánimo negativo se orientará hacia eventos frustrantes o fallidos, lo que puede impactar en la motivación y la autoeficacia [3][18].

En el contexto específico de la educación, la *Teoría del Control-Valor* de Pekrun [19] ofrece un marco para clasificar y explicar el origen de las emociones académicas. Esta teoría distingue principalmente entre *emociones de logro*, que se refieren a actividades y resultados evaluados según estándares de competencia, y *emociones epistémicas* que están directamente vinculadas al proceso de construcción de conocimiento [20]. Las emociones de logro se dividen, a su vez, según su objeto (la actividad en sí o el resultado) y su valencia (positiva o negativa). De esta forma, surgen emociones como el disfrute durante el estudio (actividad, positiva), la ansiedad ante un examen (resultado prospectivo, negativa), el orgullo por una buena calificación (resultado retrospectivo, positiva), o el aburrimiento en clase (actividad, negativa) [17][19]. Esta teoría postula que las emociones son el resultado de las valoraciones que hace el estudiante sobre dos factores clave, como lo es: el control percibido sobre la actividad o el resultado, y el valor subjetivo que le atribuye [19].

Otra teoría esencial para el ámbito educacional es la Teoría del Flujo (*Flow Theory*) de Csikszentmihalyi [22], la cual describe la “experiencia óptima” como un estado específico de equilibrio entre la habilidad del individuo y el desafío de la tarea [22][23]; es decir, es un estado de involucración total en una actividad donde nada más parece importar. Este estado representa el escenario ideal para el aprendizaje al estar asociado con una alta eficiencia cognitiva y motivación intrínseca.

La Teoría del Flujo relaciona habilidades y desafíos: el flujo se experimenta cuando las habilidades y los desafíos están equilibrados. Dentro del estado de flujo, una característica clave es la concentración total en la tarea en cuestión, lo que implica un involucramiento profundo pero sin esfuerzo, lo que facilita una alta eficiencia cognitiva [22]. Esto se alinea directamente con el disfrute (emoción de logro positiva) que ocurre durante el estudio. Cuando este equilibrio se rompe, emergen estados emocionales que impactan negativamente en el proceso de aprendizaje. Específicamente, la ansiedad surge cuando el desafío percibido supera significativamente las habilidades del estudiante,

generando un estado de frustración que dificulta el procesamiento cognitivo y obstaculiza la adquisición de conocimiento. Por el contrario, el aburrimiento emerge cuando las habilidades del estudiante exceden considerablemente el desafío de la tarea, resultando en desmotivación y desconexión del proceso educativo [23].

Según lo anteriormente mencionado, es de esperar que emociones epistémicas como la frustración surjan cuando el estudiante se enfrenta a problemas que superan su nivel de habilidad actual. Por otro lado, la concentración facilita un procesamiento cognitivo profundo y sostenido, mientras que el aburrimiento puede señalar una falta de desafío o relevancia en la tarea. La transición entre estos estados, algunos emocionales, no es estática sino dinámica, fluctuando en respuesta a los cambios en la percepción de competencia y los retos presentados durante el proceso de aprendizaje, lo que subraya la importancia de diseñar experiencias educativas que mantengan un balance óptimo para favorecer el estado de flujo y maximizar los resultados académicos.

2.2 Fluctuación de los estados emocionales

El modelo propuesto por D'Mello y Graesser [21], y adoptado por Wang en [3], describe cómo los estudiantes transitan entre diferentes estados afectivos, partiendo de un estado base de compromiso. Cuando el estudiante encuentra un obstáculo, una contradicción o un bloqueo (considerado desequilibrio cognitivo), entra en un estado de confusión [21]. Si el estudiante logra resolver el percance puede experimentar deleite y volver a un estado de compromiso, consolidando así el proceso de aprendizaje. Sin embargo, si el esfuerzo realizado resulta insuficiente y el obstáculo persiste, la confusión deriva en un estado de frustración. La persistencia de este estado emocional negativo conduce a un estado de aburrimiento, provocando así el abandono definitivo de la tarea [6].

Estas fluctuaciones emocionales no implican una inestabilidad general, sino los cambios naturales que ocurren en respuesta a eventos del proceso de aprendizaje [21]. En [3] buscan capturar estas dinámicas modelando un “estado emocional personalizado” que evoluciona continuamente, actualizándose en cada interacción mediante la incorporación de nuevas “ganancias emocionales”. Desde la perspectiva de la Teoría del Control-Valor, estas fluctuaciones reflejan la variabilidad y maleabilidad inherentes a las emociones

académicas, las cuales cambian a lo largo del tiempo y entre distintas situaciones o dominios en función de las valoraciones de control y valor que realiza el estudiante. Lo contrario a estas fluctuaciones sería una mayor persistencia o estabilidad en un único estado emocional durante un periodo prolongado, como mantener un estado constante de aburrimiento o de flujo sin interrupciones significativas.

La influencia de estas fluctuaciones emocionales en el rendimiento actúan como un factor personalizador que modula la ganancia de conocimiento (*Emotion-Boosted learning gain*) [3]. De esta manera, las emociones impactan directamente en varios mecanismos cognitivos y motivacionales, ya que emociones negativas como la ansiedad y la frustración generan pensamientos (preocupaciones) que consumen capacidad de la memoria de trabajo; mientras que emociones positivas intrínsecas, como el disfrute de la tarea, centran la atención en el material de aprendizaje, actuando además como un potente motor para la motivación [3][17].

Para este proyecto de investigación, donde se busca extender el modelo BKT, es crucial comprender cómo los estados afectivos están intrínsecamente ligados a los procesos cognitivos. El BKT tradicional infiere el dominio de una habilidad basándose únicamente en la correctitud de las respuestas del estudiante, ignorando que el rendimiento observado es una señal imperfecta del conocimiento latente. La evidencia ha demostrado que un error puede ser producto no solo de la falta de conocimiento, sino también de un estado de frustración o confusión, mientras que una alta concentración puede indicar una interacción de aprendizaje de alta calidad, incluso si la respuesta es incorrecta. Por lo tanto, integrar las emociones permitirá que la extensión del BKT modele con mayor precisión el estado real del estudiante, diferenciando entre las influencias cognitivas y afectivas en su desempeño.

Esta dualidad en la naturaleza de la concentración observable como un valor continuo pero interpretable como un estado discreto (concentrado/no concentrado), fundamenta los tres enfoques de integración emocional exploradas en este trabajo: El primer enfoque incorpora el dato emocional de forma discreta en el espacio de observaciones (sección 3.3). El segundo enfoque preserva la estructura del BKT clásico y utiliza la concentración como un modulador continuo de la predicción del rendimiento (sección 3.4). Finalmente, el tercer enfoque combina ambos mecanismos (sección 3.5).

2.3 Modelo *Bayesian Knowledge Tracing* (BKT)

Los modelos de rastreo del conocimiento (*Knowledge Tracing*, KT) se utilizan ampliamente en los Sistemas Tutores Inteligentes (ITS) para monitorear de forma dinámica la evolución del aprendizaje de los estudiantes durante el proceso de resolución de problemas [10]. La tarea principal del KT es predecir el desempeño futuro del estudiante en una determinada habilidad o concepto, modelando la evolución de su estado de conocimiento a lo largo de una secuencia de oportunidades de práctica. Los modelos de KT operan bajo el supuesto de que el conocimiento es una variable latente (no observable directamente) que puede inferirse a partir de las respuestas observables del estudiante, clasificadas generalmente como incorrectas o correctas.

Entre los diversos enfoques del KT, los modelos probabilísticos son los más consolidados, como es el caso del *Bayesian Knowledge Tracing*. Este modelo puede describirse como un Modelo Oculto de Márkov (HMM) [9], cuyos nodos ocultos hacen referencia a los estados de conocimiento latente de los estudiantes: el estado No Dominado, define la etapa inicial donde el estudiante carece de la habilidad necesaria para resolver una tarea consistentemente, y el estado Dominado, representa la adquisición del conocimiento y su capacidad de aplicación en futuras prácticas.

El HMM se define mediante un conjunto de matrices que estructuran los parámetros del modelo, como se muestra en la Fig. 2.1. Estas matrices estocásticas deben cumplir con dos restricciones matemáticas fundamentales [2]: todos sus elementos deben ser valores no negativos, y la suma de las probabilidades en el vector inicial, así como en cada una de las filas de las matrices de transición y de emisión, debe ser igual a uno.

INICIAL (π)		TRANSICIÓN (A)			OBSERVACIONES (B)		
No dominado	$1 - P(L_o)$		No dominado	Dominado		Incorrecto	Correcto
Dominado	$P(L_o)$	No dominado	$1 - P(T)$	$P(T)$	No dominado	$1 - P(G)$	$P(G)$
		Dominado	$P(F)$	$1 - P(F)$	Dominado	$P(S)$	$1 - P(S)$

Fig. 2.1. Parámetros del Modelo BKT en su forma matricial.

Los parámetros contenidos en estas matrices tienen la siguiente interpretación conceptual [9]: el parámetro base $P(L_o)$ contenido en el *vector Inicial* (π) establece la

probabilidad inicial de dominio o conocimiento de la habilidad antes de que el estudiante realice cualquier intento. En la *matriz de transición* (A), la probabilidad de transición, también conocida como parámetro de aprendizaje $P(T)$, se refiere a la probabilidad de que el estudiante adquiera el conocimiento transitando del estado no dominado al estado dominado después de una oportunidad de aprendizaje; por su parte, $P(F)$ describe la probabilidad de que el estudiante olvide el conocimiento previamente adquirido y regrese al estado no dominado. Es importante señalar que, en la formulación tradicional del BKT, se asume convencionalmente que el olvido es nulo $P(F) = 0$ una vez que la habilidad ha sido dominada. Por otro lado, en la *matriz de observaciones* (B), los parámetros de emisión o desempeño relacionan el estado de conocimiento inferido con la respuesta observable del estudiante, estos son: $P(G)$, probabilidad de que un estudiante que no ha dominado la habilidad responda, aun así, correctamente; y $P(S)$, la probabilidad de que un estudiante que sí conoce la habilidad falle y responda incorrectamente.

El modelo BKT clásico se fundamenta en la distinción entre el estado de conocimiento interno del estudiante y sus respuestas a problemas o ejercicios que son observables. Esta relación se puede representar gráficamente como un HMM, tal como se ilustra en la Fig. 2.2. En esta estructura, los círculos superiores, etiquetados como K (Knowledge), simbolizan el estado de conocimiento latente en diferentes momentos del tiempo (*Tiempo 1*, *Tiempo 2*, ..., *Tiempo N*). La notación $P(L_n)$ denota la probabilidad de que el estudiante domine la habilidad en el paso del tiempo n , basándose en su desempeño previo. Estos estados latentes se estiman a partir de las secuencias de respuestas binarias (correctas o incorrectas) a problemas, las cuales constituyen los nodos observables representados por los círculos inferiores Q (Question), a lo largo de los pasos de tiempo definidos.

La evolución del conocimiento se modela mediante las flechas horizontales entre los nodos K , que representan la transición de un estado a otro, determinada por la probabilidad de aprendizaje $P(T)$. Por su parte, la conexión entre el conocimiento latente y la respuesta observable se denomina emisión, representada por las flechas verticales desde K hacia Q , y está definida por las probabilidades de adivinar $P(G)$ y de

equivocarse $P(S)$. Finalmente, el punto de partida del modelo es la probabilidad inicial de conocimiento $P(L_0)$, asociada al primer estado K antes de cualquier interacción.

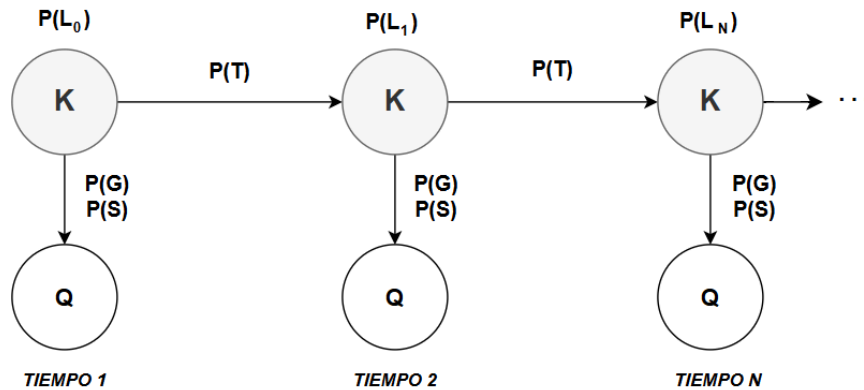


Fig. 2.2. Representación del Modelo BKT.

Específicamente, las transiciones del modelo BKT (ver Fig. 2.3), están dadas por la transición del estado “No Dominado” al estado “Dominado”. Estas se rigen por la probabilidad de transición del parámetro de aprendizaje $P(T)$, que modela la adquisición de la habilidad tras una oportunidad de práctica. No obstante, la respuesta observable del estudiante (correcta o incorrecta) se relaciona con su estado subyacente mediante los parámetros de emisión, donde $P(G)$ enlaza el estado “No Dominado” con una respuesta “Correcta”, mientras que la probabilidad de error $P(S)$ conecta el estado “Dominado” con una respuesta “Incorrecta” [11].

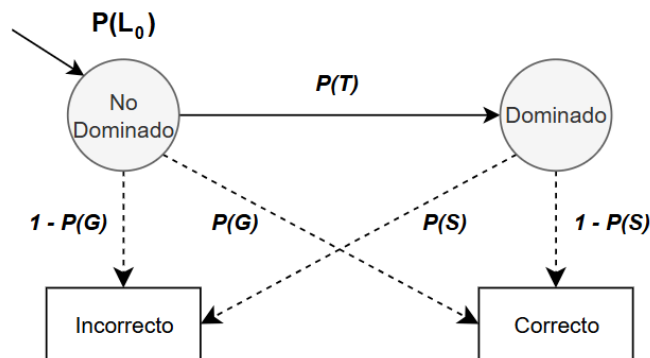


Fig. 2.3. Relación entre los estados de conocimiento y el desempeño observable.

La inferencia en el modelo BKT se realiza de manera iterativa, actualizando la probabilidad de que el estudiante domine una habilidad específica k tras cada oportunidad de práctica observada. Este proceso se formaliza mediante ecuaciones

bayesianas [2][16], donde el punto de partida del rastreo para una habilidad k es la probabilidad inicial de dominio denotada como $P(L_0)$, la cual se establece antes del primer intento del estudiante d y se define en la Ecuación 1.

$$P(L_1)_d^k = P(L_0)^k \quad (1)$$

Posterior a cada intento t , el modelo utiliza la respuesta observada para calcular la probabilidad condicional de que el estudiante domine la habilidad k . Esta inferencia bayesiana se realiza mediante dos casos mutuamente excluyentes basados en si la respuesta fue correcta o incorrecta. Si el estudiante responde correctamente ($obs = 1$), se incrementa la creencia de que domina la habilidad (ver Ecuación 2). Esta actualización calcula la probabilidad de que el estudiante realmente sepa la habilidad, dado que la respuesta fue correcta.

$$P(L_t|obs = 1)_d^k = \frac{P(L_t)_d^k \cdot (1 - P(S)^k)}{P(L_t)_d^k \cdot (1 - P(S)^k) + (1 - P(L_t)_d^k) \cdot P(G)^k} \quad (2)$$

Por lo contrario, si el estudiante responde incorrectamente ($obs = 0$), se disminuye la creencia de que domina la habilidad, (ver Ecuación 3). Este cálculo determina la probabilidad de que el estudiante realmente sepa la habilidad, a pesar de haber dado una respuesta incorrecta.

$$P(L_t|obs = 0)_d^k = \frac{P(L_t)_d^k \cdot P(S)^k}{P(L_t)_d^k \cdot P(S)^k + (1 - P(L_t)_d^k) \cdot (1 - P(G)^k)} \quad (3)$$

Una vez determinada la probabilidad condicional (Ecuaciones 2 o 3), el modelo aplica la probabilidad de transición $P(T)^k$ para modelar la adquisición de la habilidad. La probabilidad de dominio actualizada $P(L_{t+1})_d^k$ se calcula mediante la Ecuación 4, combinando la probabilidad de que el estudiante ya estuviera en el estado dominado con la probabilidad de que aprendiera la habilidad en ese paso de tiempo.

$$P(L_{t+1})_d^k = P(L_t|obs)_d^k + (1 - P(L_t|obs)_d^k) \cdot P(T)_d^k \quad (4)$$

Finalmente, para predecir el desempeño futuro del estudiante, el BKT calcula la probabilidad de que el estudiante d responda correctamente en una futura oportunidad de práctica $P(C_{t+1})_d^k$, lo cual se rige por la Ecuación 5.

$$P(C_{t+1})_d^k = P(L_t)_d^k \cdot (1 - P(S)_d^k) + (1 - P(L_t)_d^k) \cdot P(G)_d^k \quad (5)$$

La capacidad de funcionamiento del modelo depende de la estimación precisa de sus parámetros $P(L_0)$, $P(T)$, $P(G)$ y $P(S)$ para cada habilidad k . Por ello, se asume que todos estos valores están siempre entre 0 y 1 (sin incluir los extremos).

Para estimar estos valores, se implementa una variante del algoritmo de *Expectation-Maximization* (EM), conocida como el Algoritmo de *Baum-Welch*, que está específicamente diseñado para la optimización de los parámetros de los HMM [15] [16]. Mediante este procedimiento iterativo, el algoritmo encuentra el conjunto de parámetros que maximiza la verosimilitud o *likelihood* de las secuencias de respuestas observadas de los estudiantes, logrando así el mejor ajuste posible a los datos empíricos de desempeño.

2.3.1 Estimación de parámetros mediante el algoritmo *Baum-Welch*

Existen varios algoritmos que permiten realizar el ajuste de los parámetros necesarios para realizar la estimación del conocimiento latente del modelo BKT. Entre estos se encuentran métodos como Fuerza Bruta, Probabilidad Empírica, Ajuste Aleatorio, Reconocido Simulado y los basados en EM [15]. Comúnmente, se utiliza una variante del algoritmo EM, conocido como el algoritmo de *Baum-Welch*, el cual está específicamente diseñado para encontrar los parámetros de máxima verosimilitud en los HMM.

El algoritmo *Baum-Welch* es un procedimiento iterativo que ajusta los parámetros del modelo BKT ubicados en las matrices A , B y π , denotados conjuntamente como

$\theta = (A, B, \pi)$. Las matrices se mostraron en la Fig. 2.1, cuyos parámetros son probabilidad inicial del conocimiento $P(L_0)$, probabilidad de aprendizaje $P(T)$, probabilidad de adivinar $P(G)$ y probabilidad de equivocarse $P(S)$. El objetivo de este ajuste es encontrar los valores de $P(L_0)$, $P(T)$, $P(G)$, $P(S)$ que maximizan la probabilidad de observar las secuencias de respuestas dadas por los estudiantes [16]. El algoritmo alterna entre dos pasos fundamentales: el Paso de Expectativa (Paso E) y el Paso de Maximización (Paso M), repitiéndolos hasta que los valores de los parámetros alcancen la convergencia.

El Paso E tiene como objetivo calcular la distribución de probabilidad de las trayectorias de los estados latentes (los estados de conocimiento del estudiante) utilizando los parámetros del modelo estimados en la iteración actual, cabe destacar que estos pueden ser inicializados de manera aleatoria [16]. Esto se logra mediante el procedimiento Forward-Backward. La variable Forward $\alpha_i^d(t)$ definida en la Ecuación 6, representa la probabilidad de que el estudiante d se encuentre en el estado i en un tiempo t , dadas las observaciones anteriores $(y_1^d, y_2^d, \dots, y_t^d)$. En esta formulación $X_t^d = i$, X representa el estado de conocimiento del estudiante, el superíndice d identifica al estudiante, el subíndice t indica el instante de la interacción. Los valores de i corresponden a los dos estados ocultos del BKT, donde $i = 0$ hace referencia al estado “No Dominado” pero $i = 1$ al estado “Dominado”. Por consiguiente, el algoritmo utiliza esta condición para inferir el nivel de aprendizaje a partir de las observaciones y_t^d , las cuales constituyen las respuestas binarias del estudiante d en el tiempo t , siendo $y_t^d = 1$ la respuesta “Correcta” y $y_t^d = 0$ la respuesta “Incorrecta”.

$$\alpha_i^d(t) = P(y_1^d, y_2^d, \dots, y_t^d | X_t^d = i, \theta) \quad (6)$$

Para el primer paso de tiempo ($t = 1$), la Ecuación 7 se define como el producto entre la probabilidad inicial de estar en el estado i y la probabilidad de emitir la primera

observación y_1^d desde dicho estado. En esta ecuación, π_i representa el valor correspondiente al estado i dentro del vector de probabilidades iniciales π ; por su parte, $b_i(y_1^d)$ simboliza la probabilidad de observación, la cual se extrae de la matriz de observaciones B (ver Fig. 2.1). En este término, el subíndice i indica el estado latente y y_1^d representa la primera respuesta binaria (Incorrecta o Correcta) proporcionada por el estudiante d .

$$\alpha_i^d(1) = \pi_i \cdot b_i(y_1^d) \quad (7)$$

Para los siguientes pasos $(t + 1)$, la variable $\alpha_i(t + 1)$ se calcula de forma recursiva según la Ecuación 8, donde se suman las probabilidades de llegar al estado $i \in \{0, 1\}$ en el tiempo t , multiplicadas por la probabilidad de transición correspondiente a_{0i} o a_{1i} , y finalmente, multiplicando todo por la probabilidad de emitir la observación actual y_{t+1}^d . En esta fórmula, a_{ji} es la probabilidad de transitar del estado j al estado i (sacada de la matriz de transición A).

$$\alpha_i^d(t + 1) = (\alpha_0^d(t) \cdot a_{0i} + \alpha_1^d(t) \cdot a_{1i}) \cdot b_i(y_{t+1}^d) \quad (8)$$

De forma complementaria, la variable Backward $\beta_i(t)$ de la Ecuación 9, denota la probabilidad de observar la secuencia de respuestas futuras $(y_{t+1}^d, y_{t+2}^d, \dots, y_T^d)$, dado que el estudiante d se encuentra en el estado i en el tiempo t y dados los parámetros del modelo θ . Esto se calcula de forma recursiva comenzando desde el final de la secuencia y retrocediendo en el tiempo como se muestra en las Ecuaciones 10 y 11.

$$\beta_i^d(t) = P(y_{t+1}^d, y_{t+2}^d, \dots, y_T^d | X_t^d = i, \theta) \quad (9)$$

En el último instante de tiempo T , se define como:

$$\beta_i^d(T) = 1 \quad (10)$$

Ahora bien, $\beta_i^d(t)$ se calcula de forma recursiva, yendo hacia atrás en el tiempo.

Para encontrar el valor en el tiempo t , se considera cada posible estado j en el que el estudiante podría estar en el siguiente instante $t + 1$.

$$\beta_i^d(t) = \beta_0^d(t + 1) \cdot a_{i0} \cdot b_0(y_{t+1}^d) + \beta_1^d(t + 1) \cdot a_{i1} \cdot b_1(y_{t+1}^d) \quad (11)$$

La combinación de ambas variables permite calcular dos probabilidades posteriores cruciales: $\gamma_i^d(t)$, que es la probabilidad de estar en el estado i en el tiempo t dada la secuencia completa de observaciones Y y el modelo θ , expresado en la Ecuación 12.

$$\gamma_i^d(t) = P(X_t^d = i | y^d, \theta) = \frac{P(X_t^d = i, y^d | \theta)}{P(y^d | \theta)}$$

$$\gamma_i^d(t) = \frac{\alpha_i^d(t) \cdot \beta_i^d(t)}{\sum_{k=0}^1 \alpha_k^d(t) \cdot \beta_k^d(t)} \quad (12)$$

Y $\xi_{ij}^d(t)$, es la probabilidad de transitar del estado i al estado j entre el tiempo t y $t + 1$ dada la secuencia de observaciones Y y el modelo θ , expresado en la Ecuación 13.

$$\xi_{ij}^d(t) = P(X_t^d = i, X_{t+1}^d = j | y^d, \theta) = \frac{P(X_t^d = i, X_{t+1}^d = j, y^d | \theta)}{P(y^d | \theta)}$$

$$\xi_{ij}^d(t) = \frac{\alpha_i^d(t) \cdot a_{ij} \cdot b_j(y_{t+1}^d) \cdot \beta_j^d(t+1)}{\sum_{k=0}^1 \sum_{w=0}^1 \alpha_k^d(t) \cdot a_{kw} \cdot b_w(y_{t+1}^d) \cdot \beta_w^d(t+1)} \quad (13)$$

En el Paso M, el algoritmo utiliza las probabilidades esperadas de las Ecuaciones 12 y 13 calculadas en el paso E para actualizar los valores de los parámetros del modelo θ de manera que maximice la función de verosimilitud esperada, considerando "No

dominado" como el estado 0 y "Dominado" como el estado 1 [16]. La Ecuación 14 describe cómo se recalcula la probabilidad de conocimiento inicial $P(L_0)$, la cual se actualiza como la expectativa (promedio) de que un estudiante se encuentre en el estado "Dominado" en su primer intento ($t = 1$), promediada sobre todos los estudiantes, donde D representa el número total de estudiantes en el conjunto de datos utilizado para entrenar el modelo.

$$P(L_0) = \frac{1}{D} \sum_{d=1}^D \gamma_1^d(1) \quad (14)$$

La probabilidad de adivinar, se calcula nuevamente como la proporción esperada de respuestas correctas cuando el estudiante se encontraba en el estado "No dominado" (ver Ecuación 15).

$$P(G) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d \cdot \gamma_0^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)} \quad (15)$$

Análogamente, la probabilidad de desliz $P(S)$ se calcula como la proporción esperada de respuestas incorrectas cuando el estado de conocimiento era "Dominado", según Ecuación 16.

$$P(S) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} (1-y_t^d) \cdot \gamma_1^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_1^d(t)} \quad (16)$$

Finalmente, la probabilidad de aprendizaje, se actualiza como la frecuencia esperada de transiciones del estado "No dominado" al "Dominado", mostrada en la Ecuación 17.

$$P(T) = \frac{\sum_{d=1}^D \sum_{t=1}^{(T-1)^d} \xi_{01}^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^{d-1}} \gamma_0^d(t)} \quad (17)$$

Los Pasos E y M se repiten de forma iterativa, utilizando los parámetros actualizados en cada ciclo como entrada para el siguiente, hasta que los valores de los parámetros convergen. Esta convergencia se alcanza cuando el cambio en la verosimilitud entre iteraciones sucesivas es menor que un umbral predefinido, donde un posible valor es 1×10^{-4} o 0,0001. El resultado es un conjunto de parámetros que se ajusta de manera óptima a los datos de desempeño empíricos observados.

Capítulo III

Diseño

Este capítulo detalla la arquitectura de las extensiones del RE-BKT (Respuesta Emocional BKT), las cuales se dividen en tres diseños: el primer enfoque incorpora el dato emocional directamente en el nodo de observación; el segundo enfoque integra el dato emocional en la predicción; y el tercer enfoque combina los dos enfoques anteriores. A continuación, se presentan las consideraciones generales de diseño, la formulación matemática, la redefinición de los parámetros y la adaptación del algoritmo de estimación para cada variante.

3.1 Consideraciones

Para garantizar la robustez estadística en la estimación de los parámetros para cada modelo, se establecieron criterios rigurosos para la selección y el procesamiento de las secuencias de datos. En primera instancia, basándose en la investigación de Slater y Baker [4], se determinó que cada estudiante debe contar con un mínimo de 6 registros por habilidad para asegurar una convergencia fiable de los parámetros y minimizar el error en las predicciones.

Una consideración fundamental del diseño del RE-BKT, es la agrupación de los estudiantes según sus perfiles emocionales, para lo cual se adoptó una estrategia de segmentación mediante el algoritmo K-means [31], aplicado sobre un vector de trece características por estudiante que combina estadísticos descriptivos de las emociones con la tasa de aciertos y los coeficientes de correlación punto-biserial entre cada emoción y la respuesta correcta. Esta segmentación produce clústers que representan perfiles diferenciados de sensibilidad emocional. De esta forma, los parámetros estimados para cada clúster reflejan la dinámica emocional-cognitiva de cada grupo, evitando el error de especificación que surgiría de asumir parámetros homogéneos para toda la población.

En cuanto a la inicialización del algoritmo *Baum-Welch*, se reconoce que la elección de los valores iniciales es crítica debido a la tendencia del algoritmo a converger en máximos locales. Por esta razón, para los parámetros de la matriz de observación con incorporación

emocional no se define un rango restringido, sino que se mantiene aleatorio, con el propósito de permitir que el algoritmo explore libremente el espacio de parámetros y capture la influencia real de la emoción en cada perfil. Para los parámetros de conocimiento $P(L_0)$ y aprendizaje $P(T)$, en lugar de una asignación puramente aleatoria, se definieron rangos de inicialización fundamentados en [4] que favorecen la convergencia del algoritmo:

- $0,10 < P(L_0) < 0,30$ (Probabilidad inicial de conocimiento).
- $0,10 < P(T) < 0,20$ (Probabilidad de aprendizaje o transición).

Por otra parte, la detención del proceso iterativo del algoritmo se rige por dos condiciones fundamentales [26]: 1) si se supera el número máximo de iteraciones especificado o 2) si la diferencia en la log-verosimilitud (*log-likelihood*) entre iteraciones es menor que un umbral especificado. La segunda condición se cumple si la diferencia de este valor entre iteraciones sucesivas es inferior a un umbral, por esta razón el umbral de convergencia a utilizar por recomendación de esta investigación es el valor 1×10^{-4} .

Para el cálculo eficiente de este límite de convergencia, se utiliza la propiedad de la variable *Forward* (α) en el último paso de tiempo (T). De acuerdo con [27], la verosimilitud de la secuencia observada Y dado el modelo θ se puede evaluar marginalizado sobre los estados latentes finales sin necesidad de realizar la recursión de retroceso (*Backward*), aprovechando que el vector $\beta(T)$ es, por definición, un vector unitario, Ecuación 18:

$$P(Y | \theta) = \sum_i \alpha_i(T) \quad (18)$$

De esta manera, el algoritmo maximiza iterativamente la función logarítmica de esta probabilidad, garantizando que los parámetros aprendidos reflejen con precisión el comportamiento de aprendizaje observado en los datos de cada perfil estudiantil.

3.2 Extensiones RE-BKT

La investigación desarrolla tres arquitecturas del modelo RE-BKT para enriquecer el BKT tradicional mediante la incorporación de la concentración como señal emocional. Estos enfoques se diferencian por el punto del proceso en el que se integra la emoción:

1. Enfoque 1: dato emocional en el nodo de observación; este enfoque actúa en la fase de inferencia, donde se redefine la matriz de observación a un espacio discreto de 2×4 para que el conocimiento latente se actualice considerando el contexto de la respuesta y el estado emocional simultáneamente.
2. Enfoque 2: dato emocional en la predicción; este enfoque actúa en la fase de predicción donde se utiliza una función de regresión logística para que el nivel continuo de concentración sirva para modular la probabilidad de acierto (correcto), manteniendo la estructura de actualización bayesiana como en el modelo tradicional.
3. Enfoque 3: Híbrido; este enfoque tiene la particularidad de que integra ambas aproximaciones. Se utiliza la matriz extendida de 2×4 para una actualización de conocimiento más precisa y aplica la modulación logística continua para mejorar la predicción del rendimiento.

3.2.1 Premisas de diseño:

El desarrollo de estas variaciones del RE-BKT parte de las siguientes premisas:

- El rendimiento observable (correcto/incorrecto) es una señal incompleta del conocimiento latente [9][12].
- Las emociones proporcionan señales contextuales que ayudan a interpretar de manera más precisa el desempeño observado del estudiante [3][11].
- La concentración aumenta la probabilidad de éxito, mientras que estados de baja concentración pueden reducirla [3][17][21][22]. Sin embargo, la intensidad de esta relación varía entre estudiantes, lo que justifica la necesidad de determinar un perfil de sensibilidad emocional.

Partiendo de estas premisas, las modificaciones introducidas respecto al modelo BKT clásico se resumen en:

- Se incorpora la segmentación por perfil de sensibilidad emocional [31]. Para cada habilidad, se entrena una instancia por perfil emocional. Entonces, durante la predicción, el modelo selecciona los parámetros correspondientes al perfil (clúster) al que pertenece el estudiante evaluado.

- Se mantiene la estructura latente, el vector de Probabilidad Inicial (π) y la matriz de Transición (A) conservan su significado original.
- Para los enfoques 1 y 3, se expande el espacio de observaciones a 2×4 estados para incluir combinaciones de la respuesta del estudiante (Incorrecta, Correcta) y su estado de concentración (Concentrado, No concentrado).
- Para los enfoques 1 y 3 se redefine el proceso de inferencia bayesiana para incorporar el dato emocional (Concentrado, No Concentrado) en la actualización de la creencia del conocimiento.
- Para los enfoques 2 y 3 se redefine el proceso de predicción integrando una función logística que utiliza el valor continuo de la concentración para ajustar la probabilidad de respuesta correcta.

3.3 Enfoque 1: dato emocional en el nodo de observación

3.3.1 Formulación Matemática

El modelo extendido RE-BKT con el dato emocional en el nodo de observación mantiene la estructura de un BKT clásico, conservando los nodos latentes K (No dominado / Dominado) para representar el conocimiento del estudiante, pero modificando el espacio de observación. La extensión central se realiza en el nodo observable, generando cuatro observaciones posibles en cada paso del tiempo t , como se muestra en la Fig. 3.1. Este toma cuatro estados posibles en el proceso de entrenamiento: (1) Incorrecto y Concentrado, (2) Incorrecto y No Concentrado, (3) Correcto y Concentrado y (4) Correcto y No Concentrado. La categorización de estos estados se realiza mediante un umbral de concentración optimizado para cada perfil estudiantil.

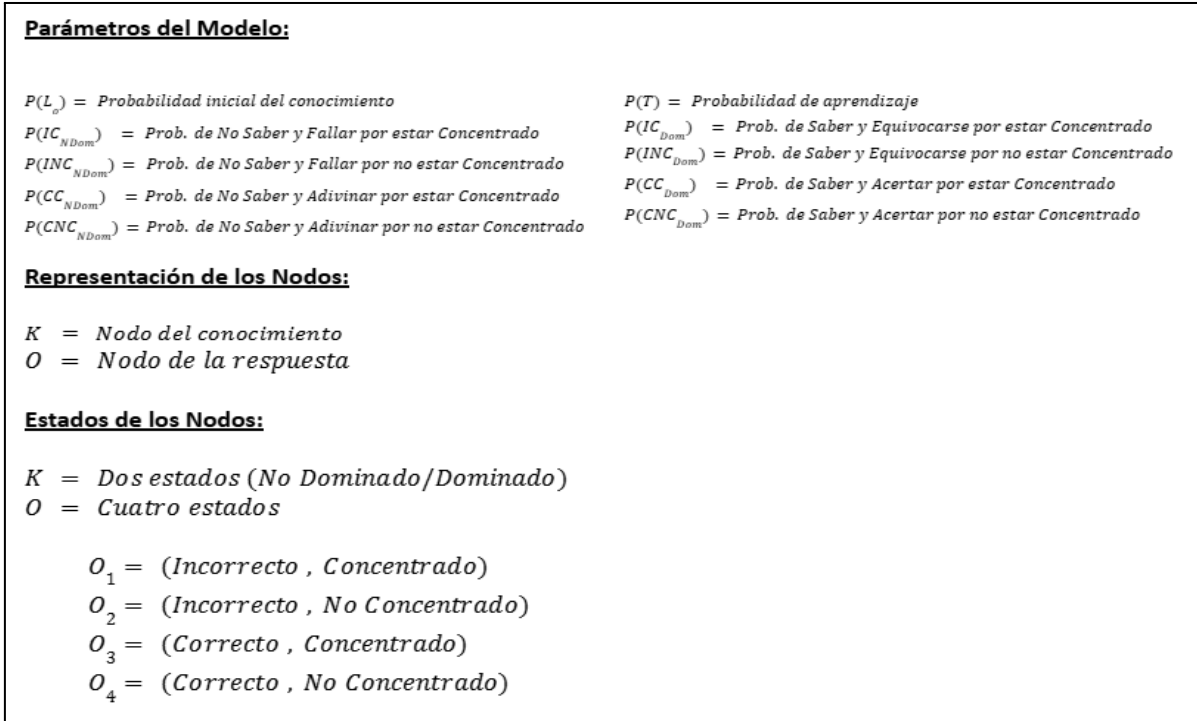


Fig. 3.1. Representación del Modelo RE-BKT con el dato emocional en el nodo de observación.

En cuanto a su representación matricial, el vector de probabilidades iniciales (π) no sufre modificaciones ya que solo describe el estado del conocimiento inicial. Así mismo, la matriz de transición (A) no se altera debido a que solo modela la transición del conocimiento entre un paso a otro, asumiendo igualmente que el olvido es nulo una vez dominada la habilidad. Por otra parte, la modificación estructural recae sobre la matriz de observaciones (B), que pasa de dimensión 2×2 en el BKT clásico a una matriz 2×4 , donde cada columna representa las observaciones compuestas generadas y cada fila corresponde a uno de los dos estados ocultos de conocimiento, como se detalla en la Tabla 3.1.

Observaciones (B)				
	Incorrecto y Concentrado	Incorrecto y No Concentrado	Correcto y Concentrado	Correcto y No Concentrado
No Dominado	$P(IC_{NDom})$	$P(INC_{NDom})$	$P(CC_{NDom})$	$P(CNC_{NDom})$
Dominado	$P(IC_{Dom})$	$P(INC_{Dom})$	$P(CC_{Dom})$	$P(CNC_{Dom})$

Tabla 3.1. Parámetros del Modelo RE-BKT con el dato emocional en el nodo de observación en su forma matricial.

En base a estas observaciones compuestas se crean parámetros propios para cada estado latente (ver Tabla 3.1). Cuando se está en el estado “No Dominado”, el estudiante no posee la habilidad, por lo tanto, lo normal es que falle ($1 - P(G)$). Si acierta, estaría adivinando ($P(G)$). Partiendo de esto, en el enfoque 1 de la arquitectura RE-BKT se definen cuatro parámetros para las cuatro observaciones posibles: $P(IC_{NDom})$ representa la probabilidad de responder incorrectamente estando concentrado, situación en la que el estudiante se esfuerza e intenta aplicar la habilidad pero falla debido a la ausencia real de conocimiento; $P(INC_{NDom})$ representa la probabilidad de responder incorrectamente por no estar concentrado, esto sucede porque el estudiante está desconectado completamente de la actividad; $P(CC_{NDom})$ representa la probabilidad de responder correctamente estando concentrado a pesar de no dominar la habilidad; y $P(CNC_{NDom})$ representa la probabilidad de responder correctamente estando no concentrado y sin dominio de la habilidad.

Por otro lado, en el estado “Dominado”, el estudiante ya posee la habilidad y lo esperado es que acierte ($1 - P(S)$). Sí llega a cometer un error se considera una equivocación o desliz ($P(S)$). En este estado se definen cuatro parámetros: $P(IC_{Dom})$ representa la probabilidad de responder incorrectamente estando concentrado, el estudiante domina la habilidad y está prestando atención, pero comete un error accidentalmente; $P(INC_{Dom})$ representa la probabilidad de responder incorrectamente estando no concentrado, esto se debe a que el estudiante domina la habilidad pero no está enfocado lo que lo lleva a cometer un error; $P(CC_{Dom})$ representa la probabilidad de responder correctamente estando concentrado, interpretado como la manifestación del aprendizaje consolidado, donde el dominio y la atención coinciden para producir un acierto genuino; y $P(CNC_{Dom})$ representa la probabilidad de responder correctamente estando no concentrado, lo que puede indicar que tuvo quizás suerte al responder ya que no estaba prestando atención.

Todos estos parámetros descritos pertenecen a la matriz de observaciones (B). Un requisito fundamental de esta matriz es que debe ser una matriz estocástica por filas, lo que significa que cada fila representa una distribución de probabilidad completa; por lo tanto, la

suma de sus elementos debe ser 1. En este enfoque, existen dos estados latentes (filas de la matriz) y cuatro observaciones posibles (columnas), por consiguiente, deben existir dos ecuaciones de restricción. La primera se asocia al estado “No Dominado”, donde la suma de las probabilidades de observar cualquiera de las cuatro combinaciones de respuesta-emoción debe ser igual a 1 como se evidencia en la Ecuación 19.

$$P(IC_{NDom}) + P(INC_{NDom}) + P(CC_{NDom}) + P(CNC_{NDom}) = 1 \quad (19)$$

La segunda restricción se asocia al estado “Dominado”, de igual manera, la suma de las probabilidades de observar cualquiera de las cuatro combinaciones debe ser 1, como se muestra en la Ecuación 20. Estas restricciones son garantizadas durante el Paso M del algoritmo *Baum-Welch* mediante la renormalización de los valores estimados tras cada iteración.

$$P(IC_{Dom}) + P(INC_{Dom}) + P(CC_{Dom}) + P(CNC_{Dom}) = 1 \quad (20)$$

Las transiciones para el modelo RE-BKT con el dato emocional en el nodo de observación se ilustran en la Fig. 3.2, donde se muestran los dos estados ocultos de conocimiento “No Dominado” y “Dominado”, con la transición entre ellos definida por la probabilidad de aprendizaje $P(T)$, y la probabilidad inicial $P(L_0)$ asociada al estado “No Dominado” antes de cualquier interacción.

A diferencia del BKT clásico, el nodo observable no representa únicamente la respuesta binaria del estudiante, sino una combinación contextualizada de respuesta y estado de concentración (concentrado o no concentrado). Las flechas que conectan los estados ocultos con este nodo observable representan las probabilidades de emisión, las cuales son parámetros únicos y fusionados que indican la probabilidad de observar la combinación completa (respuesta-emoción) dado un estado de conocimiento latente. Por ejemplo, la probabilidad de observar “Incorrecto y No Concentrado” si el estudiante “No Domina” es el parámetro $P(INC_{NDom})$, mientras que la probabilidad de esa misma observación si el estudiante “Domina” es el parámetro $P(INC_{Dom})$.

Es importante destacar que estos parámetros de emisión son estimados de forma independiente para cada combinación de habilidad y clúster. Es decir, para una misma habilidad, el modelo aprende una matriz B específica para el clúster 0 y otra distinta para el clúster 1, capturando así las diferencias en la relación entre concentración, no concentración y rendimiento que caracterizan a cada perfil estudiantil.

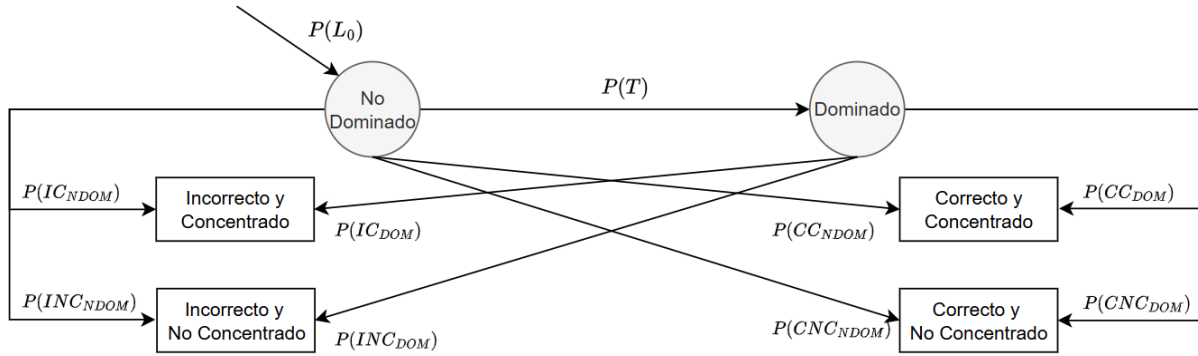


Fig. 3.2. Relación entre los estados de conocimiento y la respuesta - emoción observable.

La inferencia, es decir, la actualización de la creencia sobre el estado de conocimiento del estudiante después de cada observación, se basa en las ecuaciones bayesianas del modelo BKT, pero opera sobre el nuevo espacio de observaciones contextualizado. Inicialmente, la Ecuación 1 se mantiene idéntica debido a que la creencia inicial sobre el conocimiento no depende de ninguna observación realizada; simplemente se inicializa con el parámetro $p(L_0)^k$ para cada habilidad k y clúster correspondiente antes de cualquier interacción.

Sin embargo, las ecuaciones de actualización ya no conservan la estructura bayesiana del modelo clásico (Ecuaciones 2 y 3). Éstas se reemplazan por completo, ya que este modelo utiliza un nuevo conjunto de ecuaciones bayesianas, teniendo una para cada observación compuesta $(O_1, \dots, O_4)$. Estas ecuaciones describen la probabilidad de que el estudiante domine la habilidad después de observar la tupla (Respuesta, Emoción) en el tiempo t , haciendo uso de los ocho parámetros que influyen directamente en la inferencia del conocimiento.

Por ende, la Ecuación 21 actualiza la creencia del conocimiento cuando el estudiante falla a pesar de estar concentrado (O_1). El numerador calcula la probabilidad de la hipótesis de “Dominio”, asociada a que el estudiante ya supiera la habilidad (creencia previa $P(L_t)$) y haya

cometido una equivocación ($P(IC_{Dom})$). El denominador normaliza esta probabilidad al compararla con la hipótesis que el estudiante “No Domina” la habilidad ($1 - P(L_t)$) y cometiera un error estando concentrado ($P(IC_{NDom})$).

$$P(L_{t+1} | obs = O_1)_d^k = \frac{P(L_t)_d^k \cdot P(IC_{Dom})^k}{P(L_t)_d^k \cdot P(IC_{Dom})^k + (1 - P(L_t)_d^k) \cdot P(IC_{NDom})^k} \quad (21)$$

De manera análoga, cuando se produce la observación O_2 , que corresponde a la combinación Incorrecto y No Concentrado, la actualización bayesiana se define en la Ecuación 22. Esta inferencia estima la probabilidad de responder incorrectamente estando no concentrado ($P(INC_{Dom})$) en un estudiante que “Domina” la habilidad, frente a la probabilidad de un error por falta de concentración ($P(INC_{NDom})$) en un estudiante que “No Domina” la habilidad.

$$P(L_{t+1} | obs = O_2)_d^k = \frac{P(L_t)_d^k \cdot P(INC_{Dom})^k}{P(L_t)_d^k \cdot P(INC_{Dom})^k + (1 - P(L_t)_d^k) \cdot P(INC_{NDom})^k} \quad (22)$$

Así como se calculan las observaciones de las respuestas incorrectas desde el estado “Dominado”, se calculan las observaciones de las respuestas correctas para cada combinación. La observación O_3 que corresponde a la combinación Correcto y Concentrado, considera la evidencia más fuerte de aprendizaje ya que analiza la hipótesis de que el aprendizaje consolidado ($P(CC_{Dom})$) será significativamente mayor que la adivinanza estando concentrado ($P(CC_{NDom})$).

$$P(L_{t+1} | obs = O_3)_d^k = \frac{P(L_t)_d^k \cdot P(CC_{Dom})^k}{P(L_t)_d^k \cdot P(CC_{Dom})^k + (1 - P(L_t)_d^k) \cdot P(CC_{NDom})^k} \quad (23)$$

Finalmente, la Ecuación 24 maneja la observación O_4 (Correcto, No Concentrado), evalúa si el acierto se produjo a pesar de que el estudiante que “Domina” no está concentrado

$(P(CNC_{Dom}))$, o si fue una adivinanza de un estudiante que “No Domina” y mantiene el mismo nivel de desconcentración $(P(CNC_{NDom}))$

$$P(L_{t+1} | obs = O_4)_d^k = \frac{P(L_t)_d^k \cdot P(CNC_{Dom})^k}{P(L_t)_d^k \cdot P(CNC_{Dom})^k + (1 - P(L_t)_d^k) \cdot P(CNC_{NDom})^k} \quad (24)$$

Una vez que se ha calculado la probabilidad posterior mediante alguna de las ecuaciones anteriores, se aplica la transición de aprendizaje $P(T)$, por lo que la Ecuación 25 se mantiene formalmente igual que en el BKT clásico pero con la diferencia de que el valor de $P(L_{t+1} | obs)$ incorpora en sí, la evidencia emocional del perfil.

$$P(L_{t+1})_d^k = P(L_{t+1} | obs)_d^k + (1 - P(L_{t+1} | obs)_d^k) \cdot P(T)^k \quad (25)$$

Para predecir el desempeño futuro del estudiante, el modelo RE-BKT con el dato emocional en el nodo de observación calcula la probabilidad total de una respuesta “Correcta” (independientemente de la emoción) sumando las probabilidades de todas las observaciones futuras que implican un acierto (O_3 y O_4). La predicción se basa en la creencia de dominio actual $P(L_t)$ y los parámetros de observación fusionados, como se muestra en la Ecuación 26:

$$P(C_{t+1})_d^k = P(L_t)_d^k \cdot (P(CC_{Dom})^k + P(CNC_{Dom})^k) + (1 - P(L_t)_d^k) \cdot (P(CC_{NDom})^k + P(CNC_{NDom})^k) \quad (26)$$

Esta ecuación suma dos componentes principales: la probabilidad de que el estudiante “Domine” la habilidad y acierte (ya sea Concentrado o No Concentrado), y la probabilidad de que “No Domine” pero acierte de todos modos (una adivinanza en cualquiera de los dos estados emocionales). Esta suma es válida porque los eventos O_3 (Correcto y Concentrado) y O_4 (Correcto y No Concentrado) son mutuamente excluyentes dentro de la categoría *respuesta correcta*, por lo que la probabilidad marginal de acertar, independientemente del estado emocional, es su suma directa.

3.3.2 Estimación de parámetros

Los diez parámetros del Modelo RE-BKT con el dato emocional en el nodo de observación se definen mediante:

$\theta = (P(L_o), P(T), P(IC_{NDom}), P(INC_{NDom}), P(CC_{NDom}), P(CNC_{NDom}), P(IC_{Dom}), P(INC_{Dom}), P(CC_{Dom}), P(CNC_{Dom}))$, los cuales son estimados a partir de los datos de entrenamiento mediante el algoritmo *Baum-Welch* descrito en el Capítulo II. Este algoritmo ajustará iterativamente los parámetros para maximizar la verosimilitud de las secuencias de observación compuestas (Respuesta, Emoción) registrada de los estudiantes, tomando en cuenta que este proceso se ejecuta de forma independiente para cada combinación de habilidad y clúster. Esta estrategia garantiza que los parámetros aprendidos reflejen la dinámica emocional-cognitiva particular de cada perfil estudiantil.

Para la implementación del algoritmo, se definen las matrices de parámetros π , A , B . Se comienza con el vector de estados iniciales, el cual se define de la misma forma que en el BKT clásico, pero inicializado independientemente para cada clúster con el parámetro $P(L_0)$ estimado a partir de las secuencias de ese perfil: donde $\pi = \{\pi_i\} = \{P(X_0^d = i)\}$ para $i = 0, 1$:

$$\pi = \begin{bmatrix} X_0^d = 0 \\ X_0^d = 1 \end{bmatrix} = \begin{bmatrix} 1 - P(L_0) \\ P(L_0) \end{bmatrix}$$

La matriz de transición A describe la probabilidad de cambiar de un estado de conocimiento a otro entre pasos de tiempo. Los estados ocultos están denotados por X_t . La matriz se define como $A = \{a_{ij}\} = \{P(X_t^d = j | P(X_{t-1}^d = i))\}$ para $i = 0, 1$ y $j = 0, 1$.

$$A = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} = \begin{bmatrix} 1 - P(T) & P(T) \\ 0 & 1 \end{bmatrix}$$

La matriz de observación B se extiende a una matriz 2×4 que representa la probabilidad de una observación i dado un estado de conocimiento oculto j . Esta se define

como $B = \{b_j(i)\} = \{P(Y_t^d = i | X_t^d = j)\}$, donde $i \in \{O_1, \dots, O_4\}$ y $j \in \{0, 1\}$ son los estados ocultos (No Dominado / Dominado).

Basándonos en la Tabla 3.1, la matriz B se define como:

$$B = \begin{bmatrix} b_{00}(O_1) & b_{01}(O_2) & b_{02}(O_3) & b_{03}(O_4) \\ b_{10}(O_1) & b_{11}(O_2) & b_{12}(O_3) & b_{13}(O_4) \end{bmatrix}$$

Sustituyendo los parámetros del modelo RE-BKT:

$$B = \begin{bmatrix} P(IC_{NDom}) & P(INC_{NDom}) & P(CC_{NDom}) & P(CNC_{NDom}) \\ P(IC_{Dom}) & P(INC_{Dom}) & P(CC_{Dom}) & P(CNC_{Dom}) \end{bmatrix}$$

Teniendo en cuenta la definición de las matrices, se ejecuta el algoritmo *Baum-Welch*, donde se inicializa los parámetros con valores aleatorios (que cumplan las restricciones de probabilidad) y se itera entre el Paso E y el Paso M.

Paso E (Expectativa)

El procedimiento Forward (cálculo de α) utiliza la misma Ecuación 6 del BKT clásico, teniendo en cuenta que esta ecuación recursiva no cambia en absoluto, lo que cambia es el significado de $b_j(y_{t+1})$, el cual es la probabilidad de la observación contextualizada y_{t+1} dado el estado oculto j .

Cuando hablamos de y_{t+1} nos referimos a una de las dos observaciones, por ejemplo $y_{t+1} = O_1$ (Incorrecto, Concentrado), por ende, si el estado latente es $j = 0$ (“No Dominado”), el valor de $b_0(y_{t+1}) = P(IC_{NDom})$. Si el estado latente es 1 (“Dominado”), entonces $b_1(y_{t+1}) = P(IC_{Dom})$, obteniendo así:

$$\alpha_0^d(t+1) = \left[\alpha_0^d(t) \cdot a_{00} + \alpha_1^d(t) \cdot a_{10} \right] \cdot \left[P(IC_{NDom}) \right]$$

$$\alpha_1^d(t+1) = \left[\alpha_0^d(t) \cdot a_{01} + \alpha_1^d(t) \cdot a_{11} \right] \cdot \left[P(IC_{Dom}) \right]$$

De manera análoga, el procedimiento Backward (β) se calcula utilizando la Ecuación 9, funcionando con la misma lógica pero en sentido inverso en el tiempo.

$\gamma_i^d(t)$ y $\xi_{ij}^d(t)$ se calculan después que α y β , utilizando las mismas fórmulas que en el BKT clásico (Ecuaciones 12 y 13), incorporando ya toda la evidencia de la observación contextualizada (Respuesta, Emoción).

Paso M (Maximización)

En el paso M, los parámetros iniciales (π) y de transición (A), representados por $P(L_0)$ y $P(T)$, mantienen la lógica de actualización del modelo BKT clásico mediante las Ecuaciones 14 y 17, dado que estos parámetros describen exclusivamente la dinámica del estado de conocimiento con independencia del estado emocional. La actualización de la matriz de observaciones (B) utiliza la lógica inicial de las ecuaciones del BKT clásico:

$$\bar{b}_j(i) = \frac{\text{Número de veces en el estado } j \text{ observando el valor } i}{\text{Número de veces en el estado } j} \quad (27)$$

Para el estado “No Dominado” ($j = 0$), re-estimaremos los cuatro parámetros del modelo RE-BKT con el dato emocional en el nodo de observación utilizando la lógica de la Ecuación 27. En primer lugar, se define la Ecuación 28, la cual calcula la probabilidad de observar una respuesta incorrecta acompañada de concentración, dado que el estudiante aún no domina la habilidad.

$$P(IC_{NDom}) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_1 \cdot \gamma_0^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)} \quad (28)$$

Donde:

- $\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_1 \cdot \gamma_0^d(t)$: representa la suma de las probabilidades esperadas $\left(\gamma_0^d(t)\right)$ de que un estudiante d se encuentre en el estado “No Dominado” en el instante t , considerando únicamente aquellos momentos donde la observación registrada fue O_1 (Incorrecto, Concentrado).

- $\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)$: representa la normalización, calculada como la suma total de las probabilidades esperadas de que el estudiante se encuentre en el estado “No Dominado” a lo largo de todas las secuencias de tiempo, independientemente de la observación emitida.
- $\gamma_0^d(t)$: es la probabilidad posterior (calculada en el paso E) de que el estudiante d esté en el estado oculto 0 (“No Dominado”) en el tiempo t .

La Ecuación 29, calcula la probabilidad de observar una respuesta incorrecta estando no concentrado, dado que el estudiante aún no domina la habilidad.

$$P(INC_{NDom}) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_2 \cdot \gamma_0^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)} \quad (29)$$

Donde:

- $\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_2 \cdot \gamma_0^d(t)$: representa la suma de las probabilidades esperadas $\left(\gamma_0^d(t)\right)$ de que un estudiante d se encuentre en el estado “No Dominado” en el instante t , considerando únicamente aquellos momentos donde la observación registrada fue O_2 (Incorrecto, No Concentrado).
- $\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)$: representa la normalización, calculada como la suma total de las probabilidades esperadas de que el estudiante se encuentre en el estado “No Dominado” a lo largo de todas las secuencias de tiempo, independientemente de la observación emitida.
- $\gamma_0^d(t)$: es la probabilidad posterior (calculada en el paso E) de que el estudiante d esté en el estado oculto 0 (“No Dominado”) en el tiempo t .

La Ecuación 30 representa la probabilidad de no dominar la habilidad y responder correctamente estando concentrado:

$$P(CC_{NDom}) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_3 \cdot \gamma_0^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)} \quad (30)$$

Donde:

- $\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_3 \cdot \gamma_0^d(t)$: calcula el conteo esperado de veces que un estudiante *no sabía* $(\gamma_0^d(t))$ pero aun así produjo la observación O_3 (Correcto, Concentrado).
- $\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)$: es el término de normalización que representa la suma total de la probabilidad esperada de estar en el estado “No Dominado”.

La Ecuación 31 representa la probabilidad de no dominar la habilidad y responder correctamente estando concentrado:

$$P(CNC_{NDom}) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_4 \cdot \gamma_0^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)} \quad (31)$$

Donde:

- $\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_4 \cdot \gamma_0^d(t)$: calcula el conteo esperado de veces que un estudiante *no sabía* $(\gamma_0^d(t))$ pero aun así produjo la observación O_4 (Correcto, No Concentrado).
- $\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_0^d(t)$: es el término de normalización que representa la suma total de la probabilidad esperada de estar en el estado “No Dominado”.

De manera similar se calculan los parámetros para el estado latente “Dominado” ($j = 1$), obteniendo así las ecuaciones 32, 33, 34 y 35.

$$P(IC_{Dom}) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_1 \cdot \gamma_1^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_1^d(t)} \quad (32)$$

$$P(INC_{Dom}) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_2 \cdot \gamma_1^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_1^d(t)} \quad (33)$$

$$P(CC_{Dom}) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_3 \cdot \gamma_1^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_1^d(t)} \quad (34)$$

$$P(CNC_{Dom}) = \frac{\sum_{d=1}^D \sum_{t=1}^{T^d} y_t^d = O_4 \cdot \gamma_1^d(t)}{\sum_{d=1}^D \sum_{t=1}^{T^d} \gamma_1^d(t)} \quad (35)$$

En cada ecuación, el numerador acumula la probabilidad esperada de estar en el estado j únicamente en los instantes donde la observación registrada coincide con i , mientras que el denominador representa la normalización sobre la totalidad del tiempo esperado en el estado j . Al finalizar el Paso M, los valores actualizados de cada fila de la matriz B son renormalizados para garantizar el cumplimiento de las restricciones estocásticas de las Ecuaciones 19 y 20. Es pertinente resaltar que la clasificación de las respuestas en las categorías de observación depende del valor continuo de la concentración, el cual permite discernir, mediante el umbral optimizado, si la interacción corresponde a un estado concentrado o no concentrado.

Finalmente, los Pasos E y M se ejecutan de manera alterna e iterativa hasta que se cumple alguno de los criterios de convergencia descritos en la Sección 3.1, produciendo como

resultado los parámetros óptimos θ_k^0 y θ_k^1 para cada habilidad k y cada perfil estudiantil, respectivamente.

3.4 Enfoque 2: dato emocional en la predicción

3.4.1 Formulación Matemática

El enfoque 2 del RE-BKT mantiene la estructura de un BKT clásico, conservando los nodos latentes K (No dominado / Dominado) para representar el estado de conocimiento del estudiante. A diferencia de las otras arquitecturas, en este modelo no se altera el espacio de observaciones ni la estructura de la red bayesiana. La extensión central se realiza de manera exclusiva en la fase de predicción, donde el valor continuo de la concentración $c_t \in [0, 1]$ modula la probabilidad de acierto (respuesta correcta) mediante una función logística, ya que actúa como una señal contextual que modifica la confianza en la predicción del desempeño, sin alterar la dinámica latente del conocimiento.

En cuanto a su representación matemática, el vector de probabilidades iniciales (π) y la matriz de transición (A) y la matriz de observaciones (B) no sufren modificaciones y preservan su significado original del modelo BKT. La matriz de observaciones (B) conserva una dimensión de 2×2 y sus parámetros clásicos: la probabilidad de adivinar $P(G)$ y la probabilidad de cometer un error $P(S)$. Las observaciones generadas por el estudiante siguen evaluándose únicamente en función de su respuesta (Incorrecto/Correcto) como se puede observar en la Fig. 2.3.

En la fase de entrenamiento, se conservan las ecuaciones de actualización bayesiana del modelo clásico (Ecuaciones 2 y 3). La inicialización de la creencia $P(L_0)^k$ mediante la Ecuación 1 se mantiene idéntica, y el proceso de actualización de la probabilidad del estado de conocimiento latente $P(L_{t+1})^k$ tras cada observación, se fundamenta en la correctitud de la respuesta dada por los parámetros clásicos de emisión.

La extensión de este enfoque con respecto al modelo clásico reside en la función de predicción de la probabilidad de respuesta correcta $P(C_{t+1})$. En lugar de utilizar únicamente la creencia de dominio y los parámetros de emisión de la matriz B , el modelo incorpora el valor continuo de concentración ($c_t \in [0, 1]$) registrado en cada interacción mediante una transformación logística. Este mecanismo constituye el núcleo de la incorporación de un dato emocional en el proceso predictivo y permite que el modelo capture el efecto de la distracción sobre el rendimiento futuro, incluso en interacciones donde el estudiante no estaba concentrado y que no formaron parte del entrenamiento de la matriz B .

La predicción se realiza en dos pasos secuenciales: Primero, se calcula la *probabilidad base* de respuesta correcta a partir de la creencia de dominio actual $P(L_t)$ y los parámetros clásicos de la matriz B , utilizando la ecuación estándar del BKT (ver Ecuación 5).

$$P(C_{t+1})_{d, base}^k = P(L_t)_d^k \cdot (1 - P(S)^k) + (1 - P(L_t)_d^k) \cdot P(G)^k \quad (36)$$

Segundo, esta probabilidad base es modulada por el nivel de concentración continua del estudiante mediante una función logística. Para ello, la probabilidad base se transforma al espacio de *log-odds*, se le añade el efecto emocional ponderado y se aplica la función sigmoide para recuperar una probabilidad válida en el intervalo $[0, 1]$:

$$P(C_{t+1})_d^k = \frac{1}{1 + e^{-\text{logit}_{d, final}^k}} \quad (37)$$

$$\text{logit}_{d, final}^k = \text{logit}_{d, base}^k + (\omega \cdot (c_t - \eta)) \quad (38)$$

$$\text{logit}_{d, base}^k = \ln \left(\frac{P(C_{t+1})_{d, base}^k}{1 - P(C_{t+1})_{d, base}^k} \right) \quad (39)$$

Donde:

- ω : representa el peso emocional que controla la magnitud de la influencia de la concentración sobre la predicción.
- η : es el punto neutro que define el nivel de concentración a partir del cual la emoción tiene un efecto positivo o negativo.
- c_t : es el valor continuo de concentración observado en el instante t .

Ambos hiper parámetros ω y η son optimizados de forma independiente para cada clúster mediante el *framework Optuna* [29], utilizando como criterio de optimización el equilibrio entre la AUC, la sensibilidad y la especificidad sobre el conjunto de prueba. El espacio de búsqueda definido para la optimización comprende rangos de $\omega \in [0, 5]$ y $\eta \in [0, 1]$, de modo que la calibración refleja la sensibilidad emocional particular del perfil estudiantil correspondiente. Esta formulación garantiza que, para niveles de concentración superiores al punto neutro η , la probabilidad de respuesta correcta predicha se incrementa respecto a la estimación base del BKT; mientras que para niveles inferiores a η , la predicción se reduce, capturando el efecto negativo del desinterés o la distracción sobre el rendimiento observable.

3.4.2 Estimación de parámetros

Los cuatro parámetros base del modelo RE-BKT con incorporación del dato emocional en la predicción, denotados como $\theta = (P(L_o), P(T), P(G), P(S))$, son estimados a partir de los datos de entrenamiento mediante el algoritmo *Baum-Welch* descrito en el Capítulo II. Una característica fundamental del proceso de estimación es que este se ejecuta de forma independiente para cada combinación de habilidad y clúster, capturando así la dinámica cognitiva propia de cada perfil estudiantil sin alterar la estructura algorítmica del HMM y evitando el error de especificación que surgiría de asumir parámetros homogéneos para toda la población.

Dado que la matriz de observaciones y el espacio de estados no cambian respecto al BKT clásico, el **Paso E (Expectativa)** y el **Paso M (Maximización)** del algoritmo *Baum-Welch* se ejecutan utilizando las Ecuaciones de la 6 a la 17 descritas en la sección 2.3. El modelo entrena iterativamente sus probabilidades iniciales, de transición y de emisión

basándose netamente en la respuesta del estudiante hasta alcanzar el criterio de convergencia establecido.

Una vez entrenados los parámetros latentes θ mediante el algoritmo *Baum-Welch*, el proceso de estimación pasa a una segunda fase donde se procede a optimizar los hiperparámetros ω y η de la función logística de predicción. Mediante *Optuna*, se busca la combinación de peso emocional y punto neutro que maximice el equilibrio de las métricas de rendimiento (AUC, sensibilidad y especificidad) sobre el conjunto de prueba, para cada combinación de habilidad y clúster.

3.5 Enfoque 3: Híbrido

3.5.1 Formulación Matemática

El Enfoque 3, propone una integración estructurada de los dos enfoques descritos anteriormente. El objetivo de este modelo es aprovechar la matriz de observación extendida para inferir el estado de conocimiento latente, dependiendo del estado de la emoción (Concentrado, No Concentrado), mientras conserva la precisión del valor continuo de la concentración para modular la predicción del rendimiento futuro.

El modelo híbrido propuesto mantiene la estructura de un HMM, ya que el vector de probabilidades iniciales (π) y la matriz de transición (A) no sufren modificaciones, preservando el significado tradicional de los parámetros de conocimiento inicial $P(L_0)$ y probabilidad de aprendizaje $P(T)$. La diferencia estructural se mantiene en el manejo de las observaciones para la actualización del conocimiento, por ende, se utiliza la matriz de observaciones (B) de 2×4 definida previamente en la Tabla 3.2 de la Sección 3.3.1. Esta matriz asocia los dos estados latentes de conocimiento (Dominado/No Dominado) con las cuatro observaciones discretas compuestas: Incorrecto y Concentrado (O_1), Incorrecto y No Concentrado (O_2), Correcto y Concentrado (O_3), y Correcto y No Concentrado (O_4).

Partiendo de esta idea, el proceso de inferencia bayesiana para actualizar la creencia sobre el estado de conocimiento del estudiante, posterior a cada interacción, se rige por las

Ecuaciones 21, 22, 23 y 24 establecidas para el Enfoque 1. Una vez se calcula la probabilidad posterior $P(L_{t+1} | obs)_d^k$, se aplica la probabilidad de transición mediante la Ecuación 25.

Por otro lado, la fase de predicción incorpora la formulación matemática del Enfoque 2. En lugar de calcular la probabilidad total de una respuesta correcta tomando en cuenta sólo la suma discreta de los parámetros de la matriz B , este enfoque incorpora el valor continuo de la concentración observada ($c_t \in [0, 1]$). Esto se realiza calculando la probabilidad base de acierto mediante la Ecuación 26 y posteriormente se transforma este valor al espacio de log-odds haciendo uso de regresión logística, tal como se define en las Ecuaciones 37, 38 y 39, permitiendo así que el valor de la concentración module la predicción final de la respuesta correcta.

3.5.2 Estimación de parámetros

El proceso de estimación de parámetros de este enfoque consta de dos fases analíticas distintas, correspondientes a los componentes interno y externo del modelo. En la primera fase, se estiman los diez parámetros fundamentales que rigen la dinámica latente y discreta del modelo: $\theta = (P(L_o), P(T), P(IC_{NDom}), P(INC_{NDom}), P(CC_{NDom}), P(CNC_{NDom}), P(IC_{Dom}), P(INC_{Dom}), P(CC_{Dom}), P(CNC_{Dom}))$. Por consiguiente, se utiliza el espacio de observaciones de cuatro estados discretos (O_1, O_2, O_3, O_4). En cuanto a la aplicación del algoritmo de entrenamiento *Baum-Welch*, se mantiene matemáticamente idéntico al detallado en la Sección 3.3.2. Así mismo, el **Paso E (Expectativa)** y el **Paso M (Maximización)** utilizan iterativamente las Ecuaciones de la 28 a la 35 hasta alcanzar los criterios de convergencia, respetando la individualidad del entrenamiento para cada combinación de habilidad y clúster.

En la segunda fase, se optimizan los hiper parámetros que controlan la función de predicción logística: el peso emocional (ω), y el punto neutro (η). A diferencia de los parámetros de la matriz (B) que se ajustan mediante el algoritmo EM, los valores de ω y η se calibran mediante un enfoque de optimización bayesiana utilizando el *framework Optuna*. Este proceso busca maximizar y mantener el equilibrio de las métricas de rendimiento como la AUC, la sensibilidad y la especificidad, evaluando las predicciones del modelo sobre el

conjunto de pruebas. Al igual que en la primera fase, esta optimización se realiza de manera aislada para cada clúster, garantizando que el impacto de la concentración en la predicción refleje fielmente la sensibilidad afectiva de cada perfil de estudiantes.

Capítulo IV

Experimentación

4.1 Protocolo experimental

El diseño experimental adoptado en esta investigación responde a la necesidad de validar la eficacia de los modelos extendidos RE-BKT en comparación con el modelo estándar de *Bayesian Knowledge Tracing* implementado mediante la biblioteca pyBKT [30]. Para garantizar la reproducibilidad y la comparabilidad entre estos enfoques, el protocolo se estructuró en cinco fases secuenciales e interdependientes (ver Fig. 4.1): procesamiento de los datos, identificación de perfiles emocionales, entrenamiento mediante el algoritmo *Baum-Welch*, predicción y evaluación. Esta arquitectura experimental sigue las recomendaciones metodológicas establecidas en la literatura de minería de datos educativos [2], en la que se enfatiza la importancia de una separación estricta entre los conjuntos de entrenamiento y prueba para evitar el sobreajuste y garantizar la validez de los resultados.

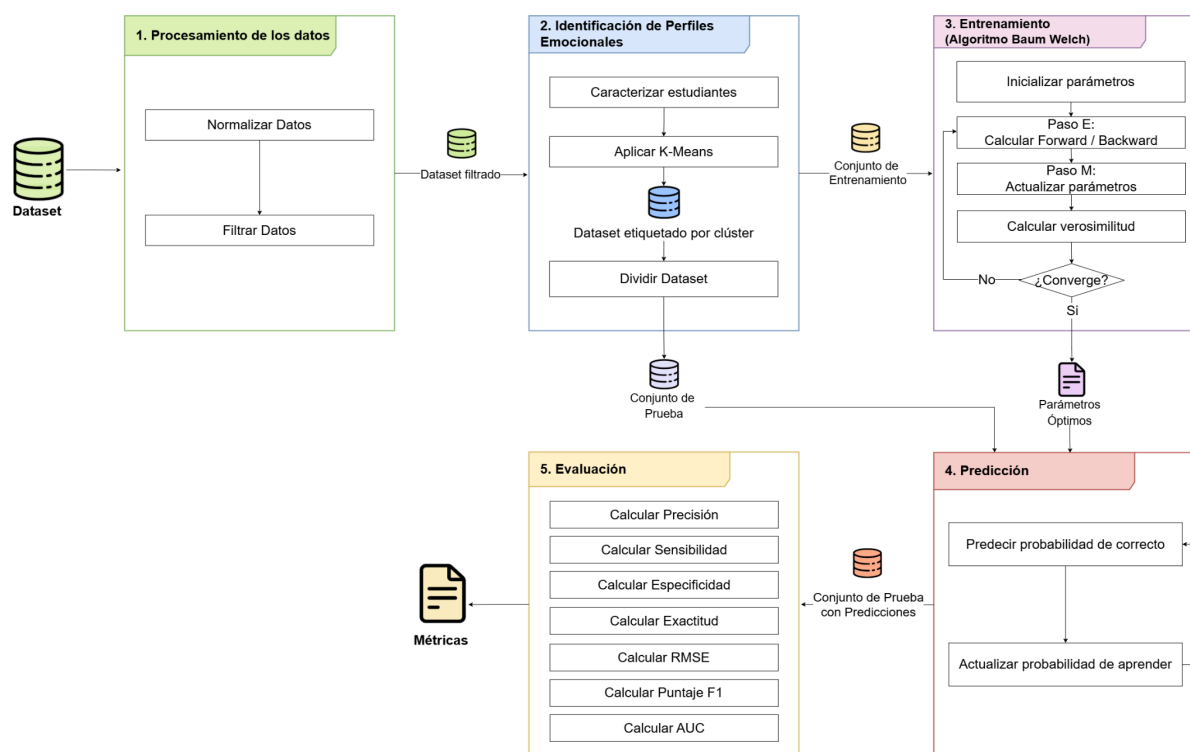


Fig. 4.1. Diagrama general de los procesos del modelo BKT.

La primera fase corresponde al **Procesamiento de los datos**, la cual comienza con el dataset que recopila las interacciones registradas entre estudiantes y un sistema tutor inteligente, incluyendo respuestas correctas e incorrectas, identificadores únicos de estudiantes, habilidades practicadas y variables emocionales continuas. Inicialmente, se llevó a cabo la normalización de los formatos de las variables para asegurar la coherencia del conjunto de datos, seguido de un protocolo de filtrado basado en un Análisis Exploratorio de Datos para garantizar la robustez estadística de la muestra, donde se excluyeron aquellos estudiantes con menos de seis interacciones registradas por habilidad. Esta decisión metodológica minimiza la influencia del ruido estocástico sobre la estimación de los parámetros del modelo [4].

En la segunda fase se integra la **Identificación de perfiles emocionales** mediante la caracterización por estudiante que combina estadísticos descriptivos de datos emocionales (concentración, frustración, confusión y aburrimiento) con la tasa de aciertos global y los coeficientes de correlación punto-biserial entre dato emocional y la respuesta correcta. Luego se aplicó el algoritmo *K-means* (como en [31]) sobre este vector multidimensional de trece características, lo que permitió identificar perfiles diferenciados de sensibilidad emocional, cuya separación fue validada mediante el coeficiente de silueta y la proyección en el plano de componentes principales. El conjunto de datos etiquetados resultante fue dividido estratégicamente en un conjunto de entrenamiento (80%) y un conjunto de prueba (20%), siguiendo el protocolo estándar ampliamente adoptado en investigaciones de seguimiento del conocimiento [9]. Esta partición garantiza que el modelo sea evaluado sobre instancias no observadas durante el entrenamiento, lo que constituye una condición indispensable para estimar su capacidad de generalización.

Con los conjuntos de datos separados se lleva a cabo la tercera fase de Entrenamiento (**Algoritmo Baum-Welch**). El proceso comienza con la inicialización de los parámetros, donde se definen los valores de partida para el vector de estados iniciales (π) y las matrices de transición (A) y observación (B). En el modelo base se establecen rangos específicos como lo sugiere la literatura previa: la probabilidad de conocimiento inicial $P(L_o)$ entre 0,10 y 0,30, la de aprendizaje $P(T)$ entre 0,10 y 0,20, la de adivinar $P(G)$ entre 0,10 y 0,20, y la de deslizamiento o error $P(S)$ entre 0,10 y 0,45. En el enfoque 1 se modifica la matriz de observación, los parámetros $P(G)$ y $P(S)$ son reemplazados por las probabilidades conjuntas

de respuesta y estado emocional (concentrado o no concentrado), estos nuevos parámetros se inicializan con valores entre 0 y 1 (sin incluir extremos), a diferencia de los rangos acotados del modelo base. En el enfoque 2 todo permanece igual al modelo base, pues la emoción se incorpora sólo en la función de predicción; y en el enfoque 3 se combinan ambas modificaciones.

A continuación, se ejecuta el Algoritmo *Baum-Welch*. Este algoritmo toma el conjunto de entrenamiento y los parámetros iniciales para iterar entre un *Paso E*, en el que calcula las probabilidades hacia adelante y hacia atrás (*Forward / Backward*) para estimar el estado oculto de conocimiento, y un *Paso M*, en el que actualiza los parámetros para maximizar la verosimilitud de los datos observados. Este ciclo se repite hasta que la diferencia entre iteraciones consecutivas es menor a 1×10^{-4} (criterio de convergencia) o hasta alcanzar un número máximo de iteraciones predefinido. Una vez cumplido el criterio, se obtienen los parámetros optimizados que mejor explican los datos de entrenamiento. Una distinción fundamental de las arquitecturas de RE-BKT respecto al modelo base es que, por cada habilidad, se entrenan instancias independientes del modelo: una optimizada para cada perfil emocional, de forma que los parámetros aprendidos reflejan las particularidades de la dinámica emocional-cognitiva de cada perfil estudiantil.

Con los parámetros estimados, se avanza a la cuarta fase de Predicción, donde el modelo utiliza los parámetros aprendidos para hacer predicciones sobre el conjunto de prueba. En el modelo base esta predicción sigue las ecuaciones estándar de BKT: se inicializa la creencia de dominio $P(L_0)$ mediante la Ecuación 1, se predice la respuesta correcta $P(C_{t+1})$ mediante la Ecuación 5, se actualiza la creencia según si la respuesta fue correcta o incorrecta mediante las Ecuaciones 2 y 3, y finalmente se aplica la transición de aprendizaje $P(T)$ para obtener $P(L_{t+1})$ mediante la Ecuación 4. En el enfoque 1, la predicción utiliza la matriz de observación extendida para calcular directamente la probabilidad de respuesta correcta. En el enfoque 2 la función de predicción incorpora la concentración como variable emocional continua mediante una transformación logística, ajustando la probabilidad de correcto en función de la distancia entre el nivel de concentración observado y un punto neutro de referencia. Y en el enfoque 3 se aplica primero la matriz de emisión extendida y sobre esta se aplica la misma modulación logística externa.

Los hiperparámetros que rigen esta modulación (el peso de la emoción, el punto neutro y el umbral de decisión) son optimizados de forma independiente para cada clúster mediante el *framework* de optimización bayesiana *Optuna* [29] en las arquitecturas que lo requieren. Esta personalización garantiza que el modelo adapte su comportamiento predictivo al perfil emocional del grupo al que pertenece el estudiante.

La quinta y última fase corresponde a la **Evaluación** del rendimiento predictivo, en la que se calculan métricas de clasificación y de error tanto para el modelo base como para los extendidos, de manera independiente por clúster para obtener una medida global del desempeño. Esta separación por perfil es fundamental, ya que permite detectar si la incorporación del dato emocional beneficia de manera diferencial a distintos tipos de estudiantes, algo que una evaluación global podría ocultar mediante el efecto de promediado.

Cabe destacar que el modelo base pyBKT requiere ser ejecutado de forma separada para cada perfil de usuario dado que su arquitectura no contempla nativamente el entrenamiento por perfil estudiantil. En los modelos extendidos RE-BKT, por el contrario, fueron diseñados con una estructura que soporta el entrenamiento diferenciado por clúster dentro de una única instancia, procesando la totalidad del conjunto de datos una vez que cada registro ha sido etiquetado con su clasificación de perfil emocional correspondiente. Esta diferencia constituye, en sí misma, una contribución metodológica relevante de la propuesta.

4.1.1 Definición de Escenarios de Prueba

Con el propósito de identificar la arquitectura óptima del modelo RE-BKT, se definieron tres escenarios (B, C, y D) de prueba que exploran distintas estrategias de integración de la información emocional en el proceso de seguimiento del conocimiento y se comparan con un modelo base (escenario A). Este enfoque iterativo de diseño experimental es coherente con la recomendación de Gonzalez-Brenes y otros [14], quienes señalan que la incorporación de características adicionales al BKT debe ser evaluada de forma progresiva para determinar qué variables aportan valor predictivo genuino y cuáles introducen ruido o complejidad innecesaria en el modelo.

Escenario A: Modelo Base - pyBKT

Este escenario corresponde al modelo base pyBKT [30], que implementa el BKT estándar sin incorporar ningún tipo de información emocional, tal como se visualiza en la Fig. 4.2, tanto la fase de entrenamiento como la fase de predicción mantienen la estructura clásica del modelo. Además utiliza un umbral de decisión fijo de 0,5 para la binarización de las probabilidades predichas y estima los cuatro parámetros clásicos del modelo $P(L_0)$, $P(T)$, $P(G)$ y $P(S)$ de forma independiente para cada habilidad. Su función dentro del protocolo experimental es la de servir como línea de referencia contra el cual se mide la mejora incremental atribuible a la incorporación de los datos emocionales. La elección de pyBKT como modelo de referencia se justifica por su amplia adopción en la comunidad de minería de datos educativos y por la disponibilidad de su implementación de código abierto [30]. Es importante resaltar que la arquitectura de este modelo no contempla la segmentación por perfil emocional, por ende, debe ejecutarse de forma independiente por cada clúster para obtener métricas comparables con el modelo extendido.

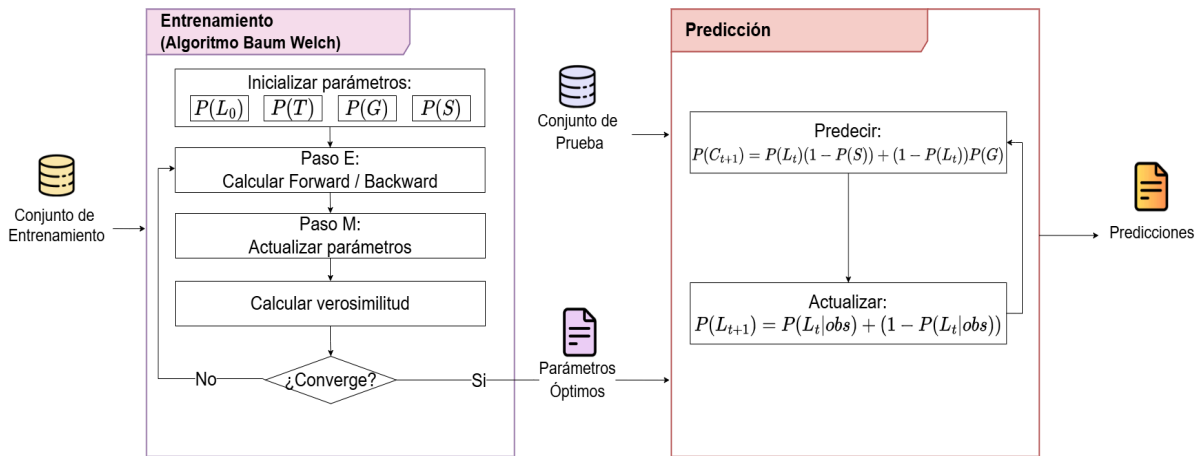


Fig. 4.2. Diagrama de las fases de entrenamiento y predicción del modelo base.

Escenario B: Enfoque 1. RE-BKT incorporando el dato emocional en el nodo de observación.

Este escenario se enfoca en la fase de entrenamiento del modelo, su característica central es la redefinición de la matriz de observación, la cual se expande de un espacio binario a un espacio discreto de 2×4 estados. Esta configuración permite que el conocimiento latente se actualice considerando simultáneamente el rendimiento observable y el estado

emocional. Dado que la variable de concentración es continua, se aplica una discretización binaria utilizando el umbral optimizado (Th), donde los valores de concentración superiores a Th se consideran “Concentrado” y los inferiores “No Concentrado”.

El modelo se entrena mediante el algoritmo *Baum-Welch*, que estima los parámetros de transición y emisión por habilidad y por clúster, sin modificar la función de predicción de respuesta correcta más allá de la matriz de emisión pero sí las ecuaciones de actualización bayesiana. La Fig 4.3 ilustra este proceso de entrenamiento en el que se inicializan los diez parámetros de observación, se llevan a cabo los pasos E y M, y se itera hasta alcanzar la convergencia. Una vez obtenidos los parámetros óptimos, se procede a la fase de predicción sobre el conjunto de prueba, donde se actualiza el conocimiento según la emoción observada. Durante la optimización con Optuna, solo se ajusta el umbral de decisión (Th) para cada clúster, ya que la emoción no introduce pesos adicionales. Este escenario permite evaluar si la inclusión directa de la concentración como un estado observable en el HMM es capaz de capturar la influencia emocional.

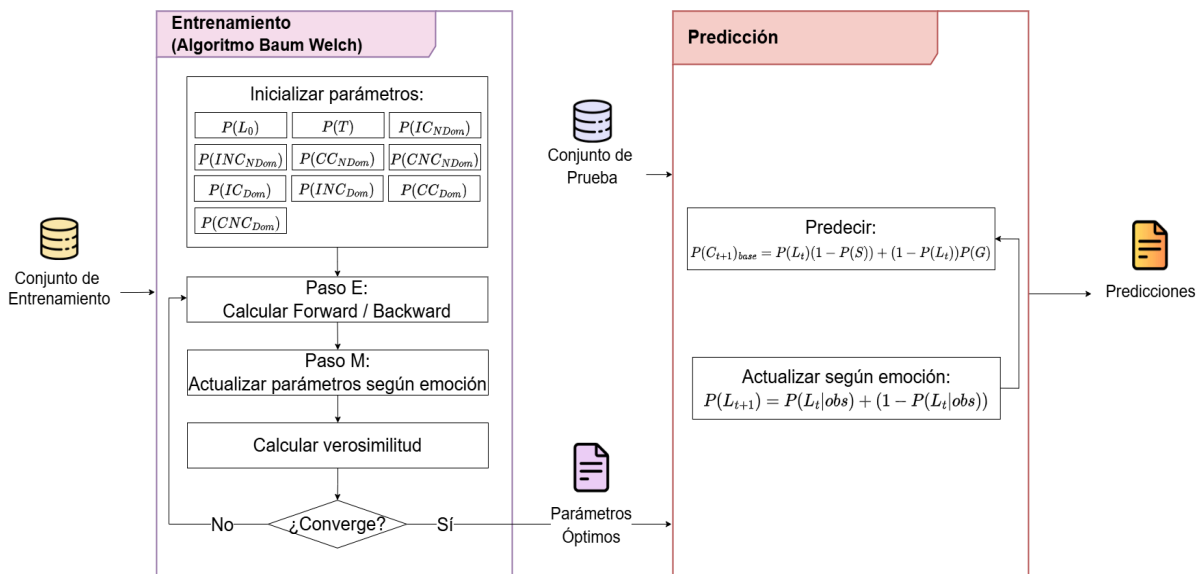


Fig. 4.3. Diagrama de las fases de entrenamiento y predicción para el Enfoque 1.

Escenario C: Enfoque 2. RE-BKT incorporando el dato emocional en la predicción.

Esta propuesta mantiene intacta la matriz de observación del BKT estándar, por lo tanto, el algoritmo *Baum-Welch* no es modificado. La incorporación de la concentración se

realiza en el cálculo de probabilidad de la respuesta correcta, mediante una regresión logística que ajusta la probabilidad obtenida del modelo base, ver Fig. 4.4. Los hiperparámetros de esta transformación (el peso emocional ω , el punto neutro np y el umbral Thm) son optimizados de forma independiente para cada clúster mediante Optuna, lo que garantiza que la personalización no se limite únicamente al proceso de entrenamiento sino que se extienda también a la fase de predicción y clasificación.

La lógica de esta arquitectura se apoya en evidencia tanto teórica como empírica que demuestra que los estudiantes no procesan ni expresan sus emociones de manera homogénea durante el aprendizaje, tal como en [21], donde mostraron que las transiciones entre estados emocionales-cognitivos (como podría ser la concentración, la confusión, la frustración o el aburrimiento) varían de un individuo a otro, ya que dependen de patrones personales vinculados con la regulación emocional.

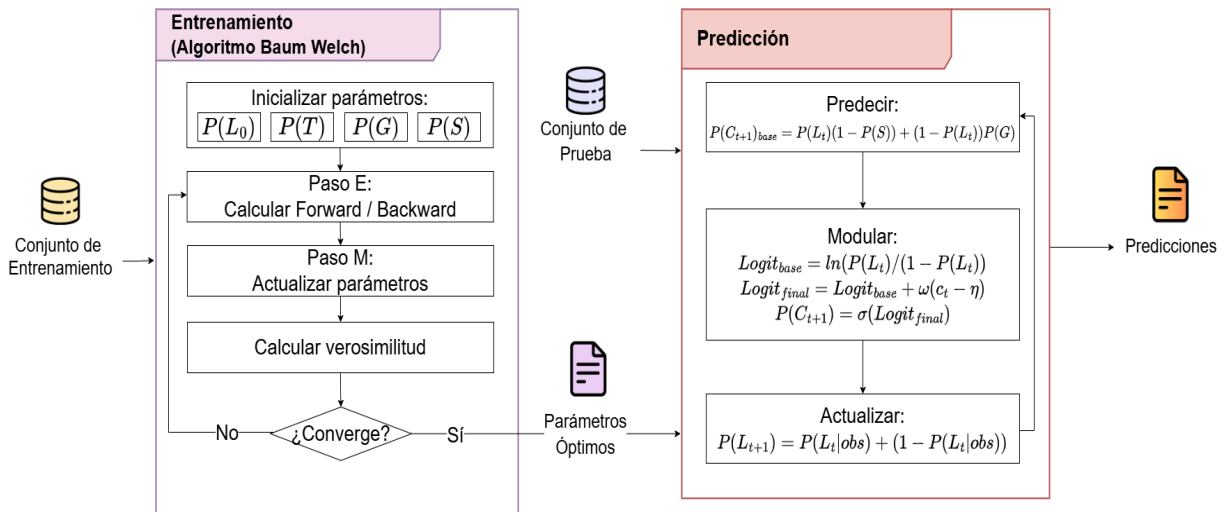


Fig. 4.4. Diagrama de las fases de entrenamiento y predicción para el Enfoque 2.

Escenario D: Enfoque 3. RE-BKT Híbrido.

Este escenario combina las dos estrategias anteriores: por un lado, la concentración se discretiza y se incluye en la matriz de emisión del modelo Baum-Welch (componente interna); por otro lado, una vez obtenidas las probabilidades del BKT extendido, se aplica una segunda modulación logística externa con los mismos hiper parámetros (ω , np y Th) que en el escenario C. Este diseño requiere un proceso de estimación dual: el ajuste de los parámetros latentes mediante el algoritmo Baum-Welch y la calibración de los hiperparámetros mediante

Optuna, garantizando que el impacto de la concentración refleje fielmente la sensibilidad emocional de cada perfil estudiantil. Este enfoque representa la máxima complejidad estructural, ya que introduce la emoción en dos puntos del flujo de inferencia como se observa en la Fig. 4.5, se modifica tanto la fase de entrenamiento como la fase de predicción.

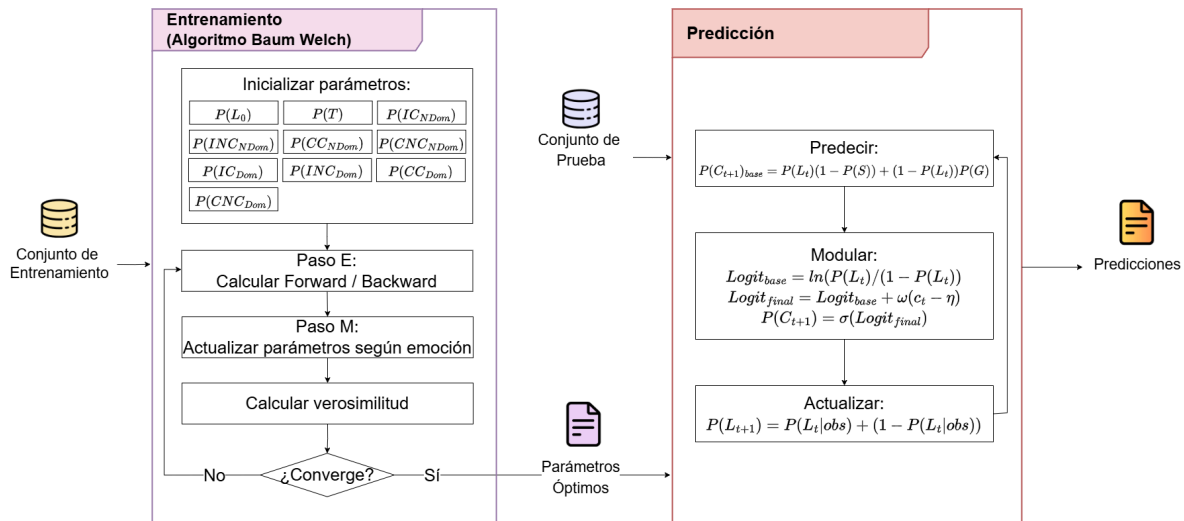


Fig. 4.5. Diagrama de las fases de entrenamiento y predicción para el Enfoque 3.

4.1.2 Conjuntos de Datos Públicos

Para el desarrollo de esta investigación, no se realizará una recolección de datos primarios, en su lugar, se utilizarán dos conjuntos de datos públicos ampliamente reconocidos en el campo de la minería de datos educativos: ASSISTments 2012-2013 [24] y ASSISTments Challenge 2017 [25]. Estos repositorios contienen registros de interacciones de estudiantes con un ITS para el aprendizaje de matemáticas, e incluyen no solo el rendimiento en cada ejercicio (respuesta correcta o incorrecta), sino también **información sobre datos emocionales del estudiante**, como la concentración, el aburrimiento, la frustración y confusión.

Como se observa en la Tabla 4.1, el dataset ASSISTments 2012-2013 [24] está compuesto por 29.018 estudiantes y 6.123.270 interacciones o registros en total, cubriendo 198 habilidades. El dataset ASSISTments Challenge 2017 [25] está compuesto por 1709 estudiantes y 942.816 interacciones, cubriendo 102 habilidades en total. Cabe señalar que las diferencias en la población estudiantil (edad, contexto escolar, entre otras), la calibración de las escalas emocionales reportadas y la dificultad de los problemas (el dataset 2017

corresponde a un “*challenge*” con tareas potencialmente más complejas) pueden influir en la relación entre emociones y rendimiento.

Estadística	Datasets	
	ASSIST2012	ASSISTchall 2017
Estudiantes	29018	1709
Interacciones	6.123.270	942.816
Habilidades	198	102

Tabla 4.1. Estadística de los datasets

Con el propósito de evaluar el comportamiento de las arquitecturas propuestas en situaciones donde no se tengan grandes conjuntos de datos con información emocional disponible, se divide cada conjunto de datos en dos subconjuntos: el **subconjunto Global** y el **subconjunto Controlado**. El **subconjunto Global** conserva la totalidad de las interacciones de los estudiantes que cumplen con el criterio mínimo de al menos seis registros por habilidad, asegurando así la robustez estadística en la estimación de los parámetros del modelo, tal como se recomienda en la literatura revisada [4]. Este subconjunto permite analizar el rendimiento de los modelos cuando se dispone de un volumen amplio de interacciones. Por otra parte, el **subconjunto Controlado** se construye a partir de los primeros seis intentos de práctica de un grupo reducido de 25 estudiantes por habilidad, buscando así simular las condiciones típicas de los entornos educativos reales, donde los datos emocionales son escasos. Esta estrategia de dividir cada dataset en una versión global y una versión controlada es consistente con metodologías previas que buscan probar la robustez de los modelos de seguimiento del conocimiento frente a la escasez de datos [4][14].

4.2 Identificación de perfiles emocionales

El análisis exploratorio de datos (*EDA*) detallado, incluyendo el preprocesamiento, el análisis de correlaciones emocionales y el procedimiento de segmentación mediante *K-means*,

se presenta en el **Apéndice A**. A continuación se resumen únicamente los perfiles de estudiantes identificados, ya que serán utilizados en las evaluaciones posteriores.

Para ASSISTments 2012-2013:

- **Estudiantes Resilientes:** estudiantes con alta tasa de aciertos (71,7 %) y baja sensibilidad emocional (correlación concentración-acierto = 0,12).
- **Estudiantes Vulnerables:** menor rendimiento (64,6 %) y mayor influencia de la concentración (correlación = 0,24) y el aburrimiento (−0,17) sobre el acierto.

Para ASSISTments 2017:

- **Estudiantes Estables de bajo rendimiento:** tasa de aciertos del 30,9 % y correlaciones moderadas (concentración = 0,30; aburrimiento = −0,32).
- **Estudiantes Concentrados pero reactivos:** mejor rendimiento (40,8 %) y alta sensibilidad a la concentración (0,48) y al aburrimiento (−0,56).

Estos perfiles se utilizaron para entrenar instancias independientes de los modelos RE-BKT por cada clúster, garantizando que los parámetros reflejan la dinámica emocional-cognitiva particular de cada grupo. Además, la concentración fue la única emoción con correlación positiva significativa en todos los perfiles, por lo que se adoptó como variable emocional principal en las arquitecturas RE-BKT.

4.3 Protocolo de Evaluación

El protocolo de evaluación se diseñó con el propósito de comparar la capacidad discriminativa y el error predictivo entre el modelo estándar pyBKT y las tres arquitecturas de la extensión RE-BKT (Enfoque 1, Enfoque 2 y Enfoque 3), asegurando que las diferencias observadas en el rendimiento sean atribuibles a la incorporación de la dimensión emocional. Para ello, se adoptó un esquema de validación basado en la división fija de los datos en un 80% para entrenamiento y un 20% para prueba, en línea con los protocolos estándar utilizados en estudios comparativos de modelos de seguimiento del conocimiento [9] [15].

Para garantizar una comparación justa, se establecieron tres condiciones de evaluación:

- **Umbrales de decisión diferenciados por clúster:** la aplicación de umbrales de decisión diferenciados por clúster para la binarización de las probabilidades predichas. Aplicar un umbral único en este contexto equivaldría a ignorar las particularidades de cada perfil, lo que comprometería la validez de la comparación.
- **Evaluación Local (por clúster):** las métricas se calculan de forma independiente para cada perfil identificado (Vulnerables/Resilientes o Reactivos/Estables), permitiendo observar cómo la sensibilidad emocional impacta en grupos específicos.
- **Evaluación Global:** se obtiene calculando la media aritmética de los resultados obtenidos en cada clúster para obtener un indicador general del desempeño para cada arquitectura del modelo RE-BKT.

4.3.1 Métricas para cada modelo

Para medir el rendimiento predictivo de los modelos evaluados, se utilizó un conjunto de métricas que abarca dos aspectos complementarios: la calidad de las predicciones binarias (Incorrecto/Correcto) y la capacidad del modelo para discriminar entre clases a partir de las probabilidades continuas que genera. Siguiendo el enfoque descrito en [32], las métricas de clasificación evalúan el comportamiento del modelo cuando se aplica una regla de decisión específica (por ejemplo, un umbral de probabilidad). En cambio, las métricas de discriminación reflejan la calidad intrínseca de la distribución de probabilidades predichas, sin depender de ningún umbral en particular. La combinación de ambos tipos de métricas ofrece una comprensión completa del desempeño para cada arquitectura del modelo RE-BKT, y permite identificar con claridad en qué aspectos la inclusión de información emocional produce mejoras reales.

Métricas de Clasificación

Estas métricas se derivan de la matriz de confusión [33], la cual estructura las predicciones del modelo en cuatro categorías: Verdaderos Positivos (VP), cuando el modelo predice correctamente que el estudiante dominará la habilidad; Verdaderos Negativos (VN), cuando predice correctamente el error; Falsos Positivos (FP), cuando predice dominio en ausencia real de él; y Falsos Negativos (FN), cuando predice error a pesar de que el estudiante efectivamente domina la habilidad, tal como se puede observar en la Tabla 4.2.

Matriz de Confusión		Predicción	
		Negativos	Positivos
Observación	Negativos	Verdaderos Negativos (VN)	Falsos Positivos (FP)
	Positivos	Falsos Negativos (FN)	Verdaderos Positivos (VP)

Tabla 4.2. Diseño de la Matriz de confusión.

Exactitud: mide la proporción global de aciertos del modelo (verdaderos positivos más verdaderos negativos) sobre el total de instancias evaluadas (Ecuación 40). Es útil como indicador general, pero debe interpretarse con precaución cuando las clases están desbalanceadas.

$$Exactitud = \frac{VP + VN}{VP + VN + FP + FN} \quad (40)$$

Precisión: cuantifica la fiabilidad de las predicciones positivas del modelo, es decir, la proporción de verdaderos positivos entre todas las instancias clasificadas como positivas. Es relevante cuando el costo de un falso positivo es alto, se describe mediante la Ecuación 41.

$$Precisión = \frac{VP}{VP + FP} \quad (41)$$

Sensibilidad: indica la capacidad del modelo para identificar correctamente los casos reales positivos (Ecuación 42), como por ejemplo el dominio de un conocimiento. Es crítica cuando no detectar un logro real lleva a subestimar el conocimiento del estudiante.

$$Sensibilidad = \frac{VP}{VP + FN} \quad (42)$$

Especificidad: Mide la capacidad del modelo para reconocer correctamente los casos negativos reales Ecuación 43, como los errores del estudiante. Útil para evitar validar respuestas incorrectas como indicadoras de dominio.

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (43)$$

Puntaje F1: media armónica entre precisión y sensibilidad, se define en la Ecuación 44. Esta métrica penaliza los desequilibrios extremos entre ambas, por lo que es el indicador más apropiado para escenarios con clases desbalanceadas.

$$\text{Puntaje F1} = \frac{\text{Precisión} \cdot \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \quad (44)$$

Métricas de Error y Discriminación

RMSE (Root Mean Squared Error): Cuantifica la magnitud promedio del error entre la probabilidad de respuesta correcta predicha por el modelo $P(C_{t+1})$ y la respuesta binaria real del estudiante y se define mediante la Ecuación 45. Un RMSE bajo indica que las probabilidades predichas están bien ajustadas a la realidad observada, lo que tiene implicaciones directas para la fiabilidad del diagnóstico del estado de conocimiento.

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (45)$$

AUC (Area Under the ROC Curve): mide la capacidad discriminativa del modelo independientemente del umbral de decisión (Ecuación 46). Representa la probabilidad de que el modelo asigna una probabilidad más alta a un caso positivo aleatorio que a uno negativo, los valores cercanos a 1 indican excelente discriminación y cercanos 0,5 es equivalente al azar. La robustez del AUC frente a los desbalances de clase y su independencia del umbral lo convierten en el estándar de referencia para la comparación de modelos de seguimiento del conocimiento en la literatura [9].

$$AUC = \frac{\sum_{l=1}^{n_1} \sum_{j=1}^{n_0} \Psi(x_l, y_j)}{n_1 \cdot n_0} \quad (46)$$

Estadístico Z de DeLong: prueba estadística no paramétrica que determina si la diferencia entre dos AUC (provenientes de modelos evaluados sobre el mismo conjunto de prueba) es significativa [28]. Calcula un estadístico Z y su valor p bajo la hipótesis nula de igualdad de áreas.

$$\text{Estadístico } Z = \frac{AUC_1 - AUC_2}{\sqrt{\text{Var}(AUC_1 - AUC_2)}} \quad (47)$$

$$p = 2 \cdot [1 - \Phi(|Z|)] \quad (48)$$

Donde Φ es la función de distribución acumulada de la normal estándar.

4.4 Análisis de Resultados

4.4.1 Comparación de métricas entre modelos

En esta sección se presenta un análisis comparativo del rendimiento predictivo entre el modelo estándar pyBKT y las tres arquitecturas diseñadas para la extensión RE-BKT (Enfoque 1, Enfoque 2 y Enfoque 3). La evaluación se realiza mediante métricas de clasificación (Precisión, Sensibilidad, Especificidad, Exactitud y F1-Score) y métricas de error y discriminación (RMSE y AUC). Los resultados, detallados en las siguientes tablas, permiten identificar cuál de las configuraciones de los modelos captura con mayor precisión la influencia de la concentración sobre el rastreo del conocimiento latente en cada perfil estudiantil identificado.

4.4.1.1 Resultados sobre el Dataset 1 (ASSISTments 2012-2013)

Como se observa en el subconjunto Global de ASSISTments 2012 (Tabla 4.3), el modelo base pyBKT presenta un sesgo marcado hacia la clase positiva, con una sensibilidad sumamente elevada (0,9509 para el clúster 0 y 0,8667 para el clúster 1) pero una especificidad crítica (0,1641 y 0,3609, respectivamente). Este comportamiento indica que el modelo tradicional tiene una alta tasa de falsos positivos. Dentro de las arquitecturas extendidas, el Enfoque 2 destaca como la configuración más equilibrada y eficiente para este dataset. Para el perfil de los “*Estudiantes Resilientes*” (clúster 0), esta arquitectura logra elevar la precisión de 0,7236 a 0,8809 y mejora drásticamente la especificidad a 0,6501. Aunque se observa una reducción en la sensibilidad, la capacidad discriminativa (AUC) asciende a 0,6806, superando al modelo base y demostrando que la modulación logística del rendimiento mediante el valor continuo de la concentración permite filtrar con mayor rigor los aciertos fortuitos.

Por otro lado, en el perfil de los “*Estudiantes Vulnerables*” (clúster 1), la arquitectura de Enfoque 2 también reporta el mejor desempeño, alcanzando un AUC de 0,7098 frente al 0,6886 de pyBKT. Resulta notable que en este grupo la precisión aumenta a 0,7762 y la especificidad se eleva a 0,6736, corrigiendo el error del modelo tradicional de subestimar las fallas en estudiantes con mayor sensibilidad emocional. Las arquitecturas de los Enfoques 1 y 3, si bien mejoran la especificidad respecto al modelo base, muestran una degradación en el AUC (situándose cerca de 0,58 - 0,66), lo que sugiere que para un volumen de datos masivo, la integración del dato emocional en la predicción consigue una mejora neta en el poder discriminativo, mientras que las variantes de los Enfoques 1 y 3 empeoran el rendimiento del modelo base. En la Tabla 4.3, el mejor resultado (Enfoque 2) se presenta en negrita, mientras que el segundo mejor rendimiento, correspondiente al modelo base pyBKT, aparece subrayado.

Modelo		ASSIST2012 - Global						
		Precisión	Sensibilidad	Especificidad	Exactitud	RMSE	F1-Score	AUC
pyBKT	<u>clúster 0</u>	<u>0,7236</u>	<u>0,9509</u>	<u>0,1641</u>	<u>0,7125</u>	<u>0,4393</u>	<u>0,8218</u>	<u>0,6735</u>
	<u>clúster 1</u>	<u>0,7090</u>	<u>0,8667</u>	<u>0,3609</u>	<u>0,6858</u>	<u>0,4533</u>	<u>0,7800</u>	<u>0,6886</u>
Enfoque 1	clúster 0	0,7443	0,5720	0,5477	0,5647	0,4603	0,6469	0,5860
	clúster 1	0,7028	0,5709	0,5662	0,5692	0,4763	0,6300	0,6002
Enfoque 2	clúster 0	0,8809	0,6115	0,6501	0,6232	0,4436	0,6935	0,6806
	clúster 1	0,7762	0,6299	0,6736	0,6455	0,4483	0,6954	0,7098
Enfoque 3	clúster 0	0,7528	0,5737	0,5664	0,5715	0,4578	0,6512	0,6040
	clúster 1	0,7348	0,7049	0,5428	0,6470	0,4630	0,7196	0,6675

Tabla 4.3. Métricas de rendimiento comparativas entre el modelo base y cada arquitectura del modelo RE-BKT para los clústers del Dataset ASSISTments 2012 (Global).

Los resultados del subconjunto Controlado se evidencian en la Tabla 4.4, donde los mejores resultados por clúster se presentan en negrita y los segundos mejores resultados subrayados. En esta tabla se observa que para el clúster 0 (*Estudiantes Resilientes*), el modelo

base alcanza una sensibilidad muy alta (0,9601) y una especificidad muy baja (0,1345), con un AUC de 0,7028 y una precisión de 0,7834 para el método base. El Enfoque 2 mejora notoriamente la especificidad (0,6849) y la precisión (0,8752), y eleva el AUC a 0,7114, superando ligeramente al modelo base. Si bien se observa una reducción en la sensibilidad, que desciende a 0,6779, esta pérdida se ve compensada por un equilibrio mucho más favorable entre sensibilidad y especificidad en comparación con el modelo base, el cual presentaba una especificidad críticamente baja. El Enfoque 1 obtiene un AUC de 0,6217 y una precisión de 0,8278, aunque con pérdida de sensibilidad. El Enfoque 3 muestra un AUC de 0,6689 y un F1-Score de 0,7546, por debajo del Enfoque 2. Por consiguiente, en el perfil *resiliente* el Enfoque 2 es la arquitectura más equilibrada, logrando el mejor compromiso entre sensibilidad y especificidad y el mayor AUC.

En lo que respecta al clúster 1 (*Estudiantes Vulnerables*), el modelo base pyBKT presenta una alta sensibilidad (0,9224) pero una baja especificidad (0,2683), reflejando nuevamente una tendencia a la sobrepredicción de aciertos. Su AUC es de 0,7105 y su precisión de 0,7771. La arquitectura del Enfoque 2 mejora sustancialmente la especificidad (0,6943 frente a 0,2683) y la precisión (0,8447), aunque la sensibilidad desciende a 0,6008. Sin embargo el AUC alcanza 0,7040, valor ligeramente inferior al del modelo base pyBKT (0,7105). El Enfoque 1, por su parte, logra un AUC de 0,6390 lo cual es inferior al base, aunque mejora la especificidad (0,6179). El Enfoque 3 obtiene un AUC de 0,6418 y una precisión de 0,8003, sin superar a la externa. En conjunto, para los *estudiantes vulnerables* del subconjunto Controlado, el Enfoque 2 no logra superar al modelo base en capacidad discriminativa general, aunque sí ofrece una mejora notable en especificidad y precisión, lo que sugiere que, en condiciones de datos muy limitados y con una correlación emoción-rendimiento débil, la ventaja de la modulación emocional externa no es concluyente y el modelo base puede resultar igualmente competitivo. En cuanto a las arquitecturas de los Enfoques 1 y 3, ambas presentaron un AUC sustancialmente más bajo que el modelo base, por lo que degradan la capacidad predictiva de forma significativa. Por lo tanto, en este escenario específico, ninguna de las dos resulta una alternativa viable al modelo clásico.

Con respecto al RMSE, todos los modelos presentan valores similares (entre 0,4094 y 0,4387), lo que indica que la inclusión de la emoción no introduce ruido adicional en la

magnitud del error cuadrático. La estabilidad del RMSE refuerza la validez de las mejoras observadas en las métricas de clasificación.

Modelo		ASSIST2012 - Controlado						
		Precisión	Sensibilidad	Especificidad	Exactitud	RMSE	F1-Score	AUC
pyBKT	clúster 0	<u>0,7834</u>	<u>0,9601</u>	<u>0,1345</u>	<u>0,7663</u>	<u>0,4094</u>	<u>0,8628</u>	<u>0,7027</u>
	clúster 1	0,7771	0,9224	0,2683	0,7487	0,4156	0,8435	0,7105
Enfoque 1	clúster 0	0,8278	0,5825	0,6050	0,5878	0,4289	0,6839	0,6217
	clúster 1	0,8145	0,6067	0,6179	0,6097	0,4325	0,6954	0,6390
Enfoque 2	clúster 0	0,8752	0,6779	0,6849	0,6795	0,4147	0,7639	0,7114
	clúster 1	<u>0,8447</u>	<u>0,6008</u>	<u>0,6943</u>	<u>0,6255</u>	<u>0,4310</u>	<u>0,7022</u>	<u>0,7040</u>
Enfoque 3	clúster 0	0,8391	0,6856	0,5714	0,6588	0,4170	0,7546	0,6689
	clúster 1	0,8003	0,6784	0,5317	0,6395	0,4387	0,7343	0,6418

Tabla 4.4. Métricas de rendimiento comparativas entre el modelo base y cada arquitectura del modelo RE-BKT para los clústers del Dataset ASSISTments 2012 (Controlado).

4.4.1.2 Resultados sobre el Dataset 2 (ASSISTments 2017)

Los resultados obtenidos sobre el subconjunto Global de ASSISTments 2017 (Tabla 4.5) revelan una dinámica predictiva distinta a la observada en el primer dataset, los estudiantes del clúster 1 presentan el mayor potencial de mejora al integrar los datos emocionales.

El modelo pyBKT tiene una especificidad elevada (0,9695 para el clúster 0 y 0,8392 para el clúster 1), pero una sensibilidad extremadamente deficiente, llegando a un valor crítico de 0,0855 en el perfil de los estudiantes estables. Esto indica que el modelo base tiene una marcada tendencia a clasificar la mayoría de las interacciones como incorrectas, subestimando los aciertos, lo que sugiere que no logra captar adecuadamente los casos de dominio de la habilidad. La arquitectura del Enfoque 2 se consolida como la solución más robusta, ya que para el perfil de los “*Estudiantes estables de bajo rendimiento*” (clúster 0) la sensibilidad se eleva de 0,0855 a 0,6408, en cuanto al AUC aumenta a 0,6689 frente al 0,6071

de pyBKT. Por su parte, en el perfil de los “*Estudiantes concentrados pero reactivos*” (clúster 1), la mejora es aún más notable ya que la sensibilidad aumenta de 0,3581 a 0,7651 y el AUC alcanza un 0,7412.

No obstante, la arquitectura del Enfoque 3 también supera al modelo base en ambos clústeres, logrando un AUC de 0,6540 en el clúster 0 y de 0,7024 en el clúster 1, situándose como la segunda mejor opción en cada caso. En contraste, la arquitectura del Enfoque 1 que presenta un rendimiento inferior al del modelo base en todas las métricas, con AUC de 0,5868 y 0,6066 respectivamente, lo que indica que la integración directa de la emoción en la matriz de observación resulta perjudicial incluso en este dataset con mayor correlación emocional. Estos incrementos demuestran que, al integrar la concentración como un modulador logístico (Enfoque 2) o mediante el modelo Híbrido (Enfoque 3), el modelo mejora la capacidad de detectar el dominio de la habilidad que el BKT clásico no logra identificar ante la complejidad de las tareas. En la Tabla 4.5, los mejores resultados por clúster se muestran en **negrita** y los segundos mejores se subrayan.

Modelo		ASSIST2017 - Global						
		Precisión	Sensibilidad	Especificidad	Exactitud	RMSE	F1-Score	AUC
pyBKT	clúster 0	0,5577	0,0855	0,9695	0,6949	0,4544	0,1483	0,6071
	clúster 1	0,6110	0,3581	0,8392	0,6402	0,4739	0,4515	0,6558
Enfoque 1	clúster 0	0,3660	0,5734	0,5524	0,5589	0,4590	0,4468	0,5868
	clúster 1	0,4811	0,6018	0,5421	0,5668	0,4870	0,5347	0,6066
Enfoque 2	clúster 0	0,4224	0,6408	0,6052	0,6162	0,4474	0,5091	0,6689
	clúster 1	0,5671	0,7651	0,5880	0,6613	0,4676	0,6514	0,7412
Enfoque 3	<u>clúster 0</u>	<u>0,4274</u>	<u>0,5512</u>	<u>0,6672</u>	<u>0,6312</u>	<u>0,4497</u>	<u>0,4814</u>	<u>0,6540</u>
	<u>clúster 1</u>	<u>0,5355</u>	<u>0,7621</u>	<u>0,5338</u>	<u>0,6282</u>	<u>0,5120</u>	<u>0,6290</u>	<u>0,7024</u>

Tabla 4.5. Métricas de rendimiento comparativas entre el modelo base y cada arquitectura del modelo RE-BKT para los clústers del Dataset ASSISTments 2017 (Global).

Los resultados del subconjunto Controlado (Tabla 4.6) ratifican la superioridad de la arquitectura externa durante la fase de adquisición inicial del conocimiento. Para el perfil de los “*Estudiantes concentrados pero reactivos*” (clúster 0), el modelo del Enfoque 2 eleva el AUC a 0,7685 superando el 0,6732 del modelo base y logrando una exactitud de 0,7231. Es importante destacar que la arquitectura Híbrida también muestra un desempeño competitivo con un AUC de 0,6962, situándose como la segunda mejor opción para este perfil, aunque sin alcanzar la eficiencia de la versión externa pura. En contraste, la arquitectura del Enfoque 1 presenta el peor rendimiento en este clúster, con un AUC de 0,5999, muy inferior al del modelo base, lo que confirma que forzar la emoción en la matriz de observación resulta perjudicial incluso en escenarios de datos limitados.

Finalmente, para el perfil de los “*Estudiantes estables de bajo rendimiento*” (clúster 1) la arquitectura del Enfoque 2 alcanza el rendimiento más alto de toda la investigación, con un AUC de 0,7848 y una sensibilidad de 0,7843. La arquitectura Híbrida también supera al modelo base en este perfil, logrando un AUC de 0,7371, lo que la consolida como la segunda mejor alternativa en ambos clústeres. Por su parte, el Enfoque 1 vuelve a obtener el peor desempeño, ya que presenta un AUC de 0,6489, inferior incluso al modelo base, lo que refuerza la conclusión de que la integración interna de la emoción no es beneficiosa. Al igual que en los experimentos previos, el RMSE se mantiene en rangos estables (entre 0,45 y 0,52), lo que confirma que las arquitecturas propuestas optimizan la clasificación y la detección de estados latentes sin introducir inestabilidad en la magnitud del error cuadrático.

Modelo		ASSIST2017 - Controlado						
		Precisión	Sensibilidad	Especificidad	Exactitud	RMSE	F1-Score	AUC
pyBKT	clúster 0	0,6250	0,4095	0,8252	0,6523	0,4717	0,4948	0,6732
	clúster 1	0,6749	0,7085	0,6044	0,6603	0,4604	0,6913	0,7239
Enfoque 1	clúster 0	0,4811	0,5496	0,5782	0,5663	0,4871	0,5131	0,5999
	clúster 1	0,6459	0,5663	0,6401	0,6005	0,4833	0,6035	0,6489
Enfoque 2	clúster 0	0,6942	0,5970	0,8129	0,7231	0,5287	0,6419	0,7685
	clúster 1	0,6997	0,7843	0,6099	0,7036	0,4525	0,7397	0,7848

Modelo		ASSIST2017 - Controlado						
		Precisión	Sensibilidad	Especificidad	Exactitud	RMSE	F1-Score	AUC
Enfoque 3	<u>clúster 0</u>	<u>0.6128</u>	<u>0.5560</u>	<u>0.7500</u>	<u>0.6694</u>	<u>0.4845</u>	<u>0.5831</u>	<u>0.6962</u>
	<u>clúster 1</u>	<u>0.7059</u>	<u>0.6825</u>	<u>0.6070</u>	<u>0.6768</u>	<u>0.4565</u>	<u>0.6939</u>	<u>0.7371</u>

Tabla 4.6. Métricas de rendimiento comparativas entre el modelo base y cada arquitectura del modelo RE-BKT para los clústers del Dataset ASSISTments 2017 (Controlado).

En definitiva, los resultados muestran que la segmentación de estudiantes según su perfil emocional y el entrenamiento de modelos RE-BKT independientes por clúster mejora la capacidad predictiva global, especialmente en términos de AUC, Sensibilidad y Especificidad. En ASSISTments 2012-2013, la personalización logra reequilibrar el sesgo del BKT estándar hacia una mayor especificidad y precisión. En ASSISTments 2017, los perfiles emocionales son más diferenciados, ya que las correlaciones entre las emociones y el rendimiento presentan valores más extremos y contrastan fuertemente entre un clúster y otro. Esta mayor separación entre perfiles hace que la señal emocional sea más informativa, por lo que las ganancias aportadas por la arquitectura del Enfoque 2 son aún más notables, con incrementos de AUC superiores a 0,1 en ambos clústeres y repercusiones en sensibilidad.

Dataset	Subconj.	Perfil	pyBKT	Enfoque 2
ASSISTments 2012	Global	Resiliente	↓AUC = 0,6735	↑AUC = 0,6806
			↓Esp. = 0,1641	↑Esp. = 0,6501
			↓Prec. = 0,7236	↑Prec. = 0,8809
		Vulnerable	↓AUC = 0,6886	↑AUC = 0,7098
			↓Esp. = 0,3609	↑Esp. = 0,6736
			↓Prec. = 0,7090	↑Prec. = 0,7762
ASSISTments 2017	Controlado	Resiliente	↓AUC = 0,7027	↑AUC = 0,7114
			↓Esp. = 0,1345	↑Esp. = 0,6849
			↓Prec. = 0,7834	↑Prec. = 0,8752
		Vulnerable	↑AUC = 0,7105	↓AUC = 0,7040
			↓Esp. = 0,2683	↑Esp. = 0,6943
			↓Prec. = 0,7771	↑Prec. = 0,8447
ASSISTments 2017	Global	Estables de bajo rendimiento	↓AUC = 0,6071	↑AUC = 0,6689
			↓Sens. = 0,0855	↑Sens. = 0,6408
			↑Prec. = 0,5577	↓Prec. = 0,4224
		Concentrados pero reactivos	↓AUC = 0,6558	↑AUC = 0,7412
			↓Sens. = 0,3581	↑Sens. = 0,7651
			↑Prec. = 0,6110	↓Prec. = 0,5671
	Controlado	Estables de bajo rendimiento	↓AUC = 0,6732	↑AUC = 0,7685
			↓Sens. = 0,4095	↑Sens. = 0,5970
			↓Prec. = 0,6250	↑Prec. = 0,6942
		Concentrados pero reactivos	↓AUC = 0,7239	↑AUC = 0,7848
			↓Sens. = 0,7085	↑Sens. = 0,7843
			↓Prec. = 0,6749	↑Esp. = 0,6099

Tabla 4.7. Arquitectura con mejor comportamiento por escenario.

La Tabla 4.7 presenta un resumen de la arquitectura con mejor comportamiento en cada dataset. Se observa que el modelo base presenta limitaciones atribuibles a la baja especificidad (sobreestimación del conocimiento) o por muy baja sensibilidad (incapacidad para detectar el dominio de la habilidad). La incorporación de la arquitectura del Enfoque 2 mejora significativamente el rendimiento en la mayoría de los escenarios, aumentando el AUC y la sensibilidad. No obstante, su beneficio depende del contexto: en el dataset ASSISTments 2012 con pocos datos y perfil vulnerable, el modelo base es suficiente; en cambio, en ASSISTments 2017, la alta correlación emocional permite mejoras incluso con datos limitados. Estos hallazgos validan las hipótesis de que los estudiantes reaccionan de manera distinta a los estímulos emocionales y de que incorporar la concentración como un dato emocional continuo permite capturar mejor la dinámica emocional-cognitiva del aprendizaje.

4.4.2 Análisis Estadístico

Para profundizar en la validación de las extensiones arquitectónicas propuestas del modelo RE-BKT y evaluar si la mejora observada es estadísticamente significativa se aplicó la prueba de DeLong (anteriormente descrita) para la comparación de curvas ROC correlacionadas. Este análisis estadístico permite determinar si las variaciones observadas en el AUC poseen una significancia real o si, por el contrario, son producto del azar estocástico durante la inferencia sobre el conjunto de pruebas.

En todos los casos se fijó un nivel de significancia $\alpha = 0,05$. La hipótesis nula establece que no existe diferencia entre el AUC del modelo RE-BKT y el del modelo base pyBKT para el mismo escenario de evaluación. Para interpretar correctamente los resultados, se debe examinar tres valores clave de cada fila de la tabla la ΔAUC , Z , y p .

A continuación se analizan los resultados obtenidos para el subconjunto **Global** del dataset ASSISTments 2012 (Tabla 4.8). Para la arquitectura del Enfoque 2 se observa una mejora estadísticamente significativa frente al modelo base. Para el clúster 0 (*estudiantes resilientes*), se observa un incremento en el AUC de +0,0071 ($Z = 17,1935, p < 0,0001$), mientras que en el clúster 1 (*estudiantes vulnerables*) la ganancia es mayor, situándose en +0,0212 ($Z = 32,5983, p < 0,0001$). Por el contrario, los Enfoques 1 y 3 presentan

diferencias negativas (ΔAUC entre $-0,0875$ y $-0,0211$) con p-valores también significativos, lo que evidencia un rendimiento inferior al del modelo base. Estos resultados confirman que, cuando se dispone de un volumen histórico amplio de interacciones, la incorporación externa de la concentración logra mejoras pequeñas, pero estadísticamente significativas, en el poder discriminativo.

ASSIST2012 - Global					
Comparación de Modelos		AUC	ΔAUC	Estadístico Z	$p - valor$
pyBKT	clúster 0	0,6735			
	clúster 1	0,6886			
Enfoque 1	clúster 0	0,5860	$-0,0875$	$-79,0799$	$< 0,0001$
	clúster 1	0,6002	$-0,0884$	$-44,6097$	$< 0,0001$
Enfoque 2	clúster 0	0,6806	$+0,0071$	17,1935	$< 0,0001$
	clúster 1	0,7098	$+0,0212$	32,5983	$< 0,0001$
Enfoque 3	clúster 0	0,6040	$-0,0695$	$-62,6812$	$< 0,0001$
	clúster 1	0,6675	$-0,0211$	$-10,2562$	$< 0,0001$

Tabla 4.8. Resumen de la prueba de DeLong para los 4 escenarios del Dataset ASSISTments 2012 - Global.

En el subconjunto controlado (Tabla 4.9), las mejoras del Enfoque 2 son marginales ($\Delta AUC = +0,0087$ en clúster 0 y $-0,0065$ en clúster 1) y de acuerdo con la prueba de DeLong, no resultan estadísticamente significativas ($p = 0,4145$ y $p = 0,3259$). Por tanto, con datos limitados a las primeras seis interacciones, la ventaja de la arquitectura externa no es concluyente. En cambio, los Enfoques 1 y 3 muestran diferencias negativas significativas ($p < 0,0001$), lo que indica que estas configuraciones perjudican la capacidad predictiva en escenarios de pocas observaciones por estudiante.

ASSIST2012 - Controlado					
Comparación de Modelos		AUC	ΔAUC	Estadístico Z	$p - valor$
pyBKT	clúster 0	0,7027			
	clúster 1	0,7105			
Enfoque 1	clúster 0	0,6217	-0,0810	-4,4406	< 0,0001
	clúster 1	0,6390	-0,0715	-6,6689	< 0,0001
Enfoque 2	clúster 0	0,7114	+0,0087	0,8159	0,4145
	clúster 1	0,7040	-0,0065	-0,9823	0,3259
Enfoque 3	clúster 0	0,6689	-0,0338	-2,0888	0,0367
	clúster 1	0,6418	-0,0687	-5,9229	< 0,0001

Tabla 4.9. Resumen de la prueba de DeLong para los 4 escenarios del Dataset ASSISTments 2012 - Controlado.

En contraposición, en el dataset ASSISTments 2017, el subconjunto Global (Tabla 4.10), demuestra un impacto drástico y altamente significativo a favor de la personalización emocional. El Enfoque 2 produce las mejoras más notables ya que el $\Delta AUC = +0,0618$ en clúster 0 y $+0,0854$ en clúster 1, ambas altamente significativas ($p < 0,0001$), así como la arquitectura Híbrida también logra incrementos positivos ($\Delta AUC \approx +0,046$). No obstante, el Enfoque 1 empeora el AUC en comparación con pyBKT con diferencias estadísticamente significativas. Estos resultados refuerzan que, especialmente cuando los perfiles emocionales son más extremos como en este dataset, la modulación externa de la concentración aporta ganancias sustanciales y fiables.

ASSIST2017 - Global					
Comparación de Modelos		AUC	ΔAUC	Estadístico Z	$p - valor$
pyBKT	clúster 0	0,6071			
	clúster 1	0,6558			

ASSIST2017 - Global					
Comparación de Modelos		AUC	ΔAUC	Estadístico Z	$p - valor$
Enfoque 1	clúster 0	0,5868	-0,0203	-16,6342	< 0,0001
	clúster 1	0,6066	-0,0492	-27,9143	< 0,0001
Enfoque 2	clúster 0	0,6689	+0,0618	64,3817	< 0,0001
	clúster 1	0,7412	+0,0854	82,1002	< 0,0001
Enfoque 3	clúster 0	0,6540	+0,0469	26,5160	< 0,0001
	clúster 1	0,7024	+0,0466	22,1990	< 0,0001

Tabla 4.10. Resumen de la prueba de DeLong para los 4 escenarios del Dataset ASSISTments 2017 - Global.

Finalmente, la Tabla 4.11 detalla la prueba de DeLong para el subconjunto Controlado de ASSISTments 2017 donde la arquitectura del Enfoque 2 vuelve a destacar con las mayores diferencias positivas, $\Delta AUC = +0,0953$ en clúster 0 y $+0,0609$ en clúster 1, ambas con $p < 0,0001$. La arquitectura Híbrida presenta mejoras modestas ($\Delta AUC = +0,0230$ y $+0,0132$), pero en este caso los p-valores (0,0979 y 0,4060) superan el umbral del 5%, por lo que dichas diferencias no son estadísticamente significativas. El Enfoque 1 sigue mostrando un desempeño inferior (ΔAUC negativos y $p < 0,0001$), por ende, la arquitectura externa se consolida como la única que proporciona mejoras significativas y sustanciales en condiciones de datos limitados, especialmente en el perfil de estudiantes reactivos (clúster 1).

ASSIST2017 - Controlado					
Comparación de Modelos		AUC	ΔAUC	Estadístico Z	$p - valor$
pyBKT	clúster 0	0,6732			
	clúster 1	0,7239			
Enfoque 1	clúster 0	0,5999	-0,0733	-5,6453	< 0,0001
	clúster 1	0,6489	-0,0750	-4,7615	< 0,0001

ASSIST2017 - Controlado					
Comparación de Modelos		AUC	ΔAUC	Estadístico Z	$p - valor$
Enfoque 2	clúster 0	0,7685	+0,0953	9,8786	< 0,0001
	clúster 1	0,7848	+0,0609	6,6393	< 0,0001
Enfoque 3	clúster 0	0,6962	+0,0230	1,6550	0,0979
	clúster 1	0,7371	+0,0132	0,8309	0,4060

Tabla 4.11. Resumen de la prueba de DeLong para los 4 escenarios del Dataset ASSISTments 2017 - Controlado

4.4.3 Impacto de la Optimización con Optuna

La optimización de hiperparámetros en el modelo RE-BKT se realizó con el *framework* Optuna [29], una biblioteca de optimización bayesiana que explora eficientemente el espacio de búsqueda sin necesidad de realizar una costosa búsqueda en cuadrícula. Se optimizaron tres parámetros de forma independiente para cada clúster, cada conjunto de datos y para cada arquitectura de los Enfoques 2 y 3, para la arquitectura del Enfoque 1 se optimizó únicamente el umbral de decisión.

- Peso emocional (ω): controla cuánto influye la concentración en la probabilidad de respuesta correcta.
- Punto neutro (η): nivel de concentración a partir del cual la emoción tiene efecto positivo o negativo.
- Umbral de decisión (Thm): valor de probabilidad a partir del cual el modelo clasifica una respuesta como correcta (en lugar del valor fijo 0,5).

La necesidad de optimizar estos tres valores de forma separada es porque la relación entre la dimensión emocional y el desempeño cognitivo no es uniforme y varía significativamente según el perfil del estudiante. Esto se evidencia en los resultados detallados en la Tabla 4.12, que demuestran que no existe una configuración universal para el RE-BKT.

En ASSISTments 2012, la arquitectura del Enfoque 2 para el clúster 0 (*resilientes*) requiere un peso emocional de 1,1386 en el conjunto global y de 1,4364 en el controlado, con puntos neutros de 0,9811 y 0,9507 respectivamente, mientras que para el clúster 1 (*vulnerables*) los pesos son de 0,9389 y 2,0726 para cada uno. Los umbrales en este dataset son relativamente altos (0,6419 - 0,7631), lo que hace al modelo más conservador, priorizando la reducción de falsos positivos. En ASSISTments 2017, la arquitectura del Enfoque 2 alcanza pesos emocionales superiores en el subconjunto controlado (2,4558 para clúster 0 y 2,3129 para clúster 1), reflejando que la concentración tiene mayor poder predictivo en ese contexto, y los umbrales varían desde 0,3535 hasta 0,7998. La arquitectura del Enfoque 3 muestra pesos intermedios y puntos neutros que tienden a valores más bajos en ASSISTments 2017 (0,2912–0,6478), indicando una influencia diferente de la emoción.

Modelo		ASSIST2012 - Global			ASSIST2012 - Controlado		
		Peso emocional	Punto neutro	Umbral	Peso emocional	Punto neutro	Umbral
Enfoque 1	clúster 0			0,7081			0,7631
	clúster 1			0,7081			0,6972
Enfoque 2	clúster 0	1,1386	0,9811	0,6419	1,4364	0,9507	0,7320
	clúster 1	0,9389	0,5325	0,6679	2,0726	0,7950	0,7049
Enfoque 3	clúster 0	2,8860	0,6882	0,7297	1,9848	0,6517	0,7940
	clúster 1	2,8430	0,6902	0,5871	2,9900	0,5438	0,7935
		ASSIST2017 - Global			ASSIST2017 - Controlado		
Enfoque 1	clúster 0			0,3373			0,4489
	clúster 1			0,4746			0,5490
Enfoque 2	clúster 0	1,019	0,4733	0,3535	2,4558	0,6420	0,7998
	clúster 1	1,3855	0,2725	0,5000	2,3129	0,4009	0,6135
Enfoque 3	clúster 0	1,2897	0,5849	0,3561	1,5799	0,2912	0,6119
	clúster 1	1,5799	0,2912	0,6119	2,2352	0,6478	0,5433

Tabla 4.12. Resultados de la optimización de hiperparámetros por arquitectura, clúster y conjunto de datos

Nota: El conjunto Global comprende estudiantes con al menos 6 interacciones totales; el conjunto Controlado se restringe a 25 estudiantes por habilidad con exactamente las primeras 6 interacciones.

El impacto principal de *Optuna* es mejorar el equilibrio entre sensibilidad y especificidad de forma adaptativa, ya que al ajustar el umbral por clúster, el modelo puede priorizar la reducción de falsos positivos o la detección de aciertos reales. Así, la optimización permite que cada perfil estudiantil tenga su propia configuración, maximizando el rendimiento predictivo en cada escenario.

4.5 Discusión de los resultados

Antes de consolidar la propuesta final basada en perfiles de sensibilidad emocional, se exploraron distintas estrategias para integrar la señal emocional en la lógica de inferencia del BKT. Esta fase fue determinante para comprender las limitaciones de incorporar el valor de la emoción de forma directa al modelo sin hacer uso de una personalización previa del estudiante.

Inicialmente, el modelo RE-BKT extendía la matriz de observación a una dimensión de 2×6 , lo que proporcionaba doce parámetros que representaban las seis combinaciones de respuesta y datos emocionales (concentración, frustración, confusión, aburrimiento), de modo que, el algoritmo *Baum-Welch* entrenaba los parámetros por habilidad una sola vez sin ninguna distinción de perfil estudiantil. Dado que los datos emocionales en los conjuntos de datos son continuas (cada estudiante presenta un valor numérico en cada una de las columnas de concentración, frustración, confusión y aburrimiento), fue necesario discretizarlas para identificar la emoción predominante en cada interacción y construir las tuplas (Respuesta, Emoción) requeridas por el modelo. Para ello se investigaron y aplicaron tres técnicas de discretización: *Argumento Máximo*, *Ponderación por Impacto* y *Top Two Gap*.

La primera técnica aplicada fue el *Argumento Máximo*, que consiste en asignar el valor 1 al dato emocional con mayor valor continuo y 0 a las restantes, seleccionando así el dato emocional dominante en cada instante. Al entrenar el modelo con estas secuencias, los resultados obtenidos mostraron un desempeño notablemente inferior al modelo base en todas las métricas evaluadas, con un AUC de 0,5612 frente a 0,6796 del pyBKT, evidenciando que la discretización por argumento máximo elimina información emocional relevante y degrada la capacidad predictiva del modelo.

La segunda técnica para procesar los datos emocionales fue aplicar *Ponderación por Impacto*, se estudiaron cuáles podrían ser los pesos pedagógicamente coherente de las emociones y se hicieron varias pruebas. Los pesos con mayor influencia positiva sobre el modelo fueron: Concentración: 1,1; Frustración: 0,7 y Aburrimiento: 0,3. Esta aproximación fue la que tuvo mejor desempeño entre los tres experimentos, alcanzando un AUC de 0,6361, aunque sin sobrepasar al modelo base pyBKT con un AUC de 0,6796.

La tercera técnica evaluada fue *Top Two Gap* con *Ponderación por Impacto*, que considera la diferencia entre los dos valores emocionales más altos de cada instancia. Cuando esta diferencia supera un umbral, en este caso definido como $\varepsilon = 0,10$, la emoción dominante era seleccionada; en caso contrario, se aplicaba la ponderación por impacto para no descartar registros de los estudiantes procesados. Esta estrategia obtuvo un AUC de 0,5759, resultado inferior incluso al de la ponderación por impacto previa.

A pesar de implementar distintas técnicas para el tratamiento de los datos emocionales, ninguno de estos tres experimentos logró superar al modelo base en AUC y todos presentaron una especificidad baja, lo que indica un sesgo pronunciado hacia la predicción de respuestas correctas. Este comportamiento se explica en parte por el desbalance de clases presente en los conjuntos de datos: el Dataset 1 (ASSISTments 2012) presenta un 69,63% de respuestas correctas frente a un 30,37% de incorrectas, condición que favorece a que el modelo aprenda a predecir sistemáticamente la clase mayoritaria. Caso contrario del Dataset 2, en el cual el sesgo se presentó en la gran cantidad de respuestas incorrectas (62,73%) frente a las correctas (37,27%). En consecuencia, incorporar el dato emocional directamente al modelo, sin considerar la forma en que cada estudiante reacciona emocionalmente durante el aprendizaje, introduce ruido en lugar de información útil.

Esta razón motivó la exploración de la caracterización del estudiante mediante perfiles de sensibilidad emocional, dando origen a la estrategia de segmentación por clúster explorando distintas arquitecturas. Por lo tanto, la decisión de incorporar el clústering no fue una decisión empírica ni arbitraria, si no una elección fundamentada en la evidencia experimental acumulada durante la exploración.

Ahora bien, los resultados de la evaluación comparativa entre el modelo base pyBKT y las tres arquitecturas de RE-BKT, se destaca el modelo del **Enfoque 2** el cual incorpora la concentración mediante una regresión logística posterior al paso BKT. Ya que este mejora

consistentemente el AUC respecto al modelo base, alcanzando incrementos estadísticamente significativos ($p < 0,0001$) en la mayoría de los escenarios, con ΔAUC que van desde +0,0071 en ASSIST2012 Global hasta +0,0953 en ASSIST2017 Controlado. Por el contrario, la arquitectura del Enfoque 1 degrada el AUC en todos los escenarios evaluados y la arquitectura Híbrida tuvo un comportamiento mixto: en ASSIST2012 también degradó el AUC respecto al modelo base; sin embargo, en ASSIST2017 logró mejoras pequeñas pero consistentes en el AUC para ambos subconjuntos, esto se le puede atribuir a que la modulación externa trata de mejorar la arquitectura pero al expandir la matriz de observación resulta perjudicial.

La superioridad de la arquitectura del Enfoque 2 se explica porque la concentración actúa como un modulador de la confianza en la predicción, no como un estado oculto adicional; al ajustar la probabilidad de acierto a posteriori, se corrige el sesgo del modelo sin alterar la estructura fundamental del BKT. Además, la segmentación por clústeres es imprescindible ya que los hiperparámetros óptimos (peso emocional, punto neutro, umbral) difieren marcadamente entre perfiles vulnerables y resilientes, y también entre conjuntos de datos. Por esto, la idea de que una respuesta correcta producida en un estado de alta concentración es más informativa sobre el dominio real de la habilidad que una producida bajo aburrimiento o desinterés, en línea con los hallazgos de [3].

En ASSISTments 2012, las mejoras de la arquitectura del Enfoque 2 en el conjunto Global son significativas mientras que en el Controlado las diferencias pierden significación estadística. Esto indica que, cuando los datos son escasos, la ventaja del Enfoque 2 se reduce y el modelo base es tan bueno como el RE-BKT; en cambio, en ASSISTments 2017, en ambos subconjuntos la arquitectura externa logra mejoras notables y significativas. A diferencia de los modelos basados en redes neuronales descritos en [3][11], que requieren volúmenes masivos de registros para evitar el sobreajuste, la arquitectura propuesta preserva la eficiencia paramétrica del BKT clásico, demostrando que la señal emocional es clave y que incluso en escenarios con alta reactividad emocional muestras pequeñas pueden ser suficientes y esto va a depender de la fuerza de la asociación emoción-rendimiento.

En última instancia, cualquier estrategia de integración emocional que asuma parámetros homogéneos para toda la población cae en un error de especificación que degrada su capacidad predictiva en subgrupos con dinámicas emocional-cognitivas diferenciadas. La

personalización por clúster, tal como se implementa en el RE-BKT, constituye una respuesta metodológicamente fundamentada a esta limitación estructural del modelo clásico.

4.6 Comparación con trabajos previos

La tabla 4.13 presenta una comparación entre los modelos de seguimiento del conocimiento propuestos en la literatura y nuestra propuesta considerando los siguientes criterios: las variables del contexto que usan, la arquitectura base de la propuesta, si usan datos emocionales, si requieren grandes grupos de datos para entrenarse, y si el modelo es interpretable. La tabla 4.13 permite observar claramente las diferencias arquitectónicas y operativas, agrupándolos en tres familias. El BKT clásico [30] y el FAST [14] representan enfoques interpretables pero limitados: el primero, basado en un HMM, ignora por completo cualquier información emocional, mientras que FAST utiliza regresión logística, requiere una cuidadosa ingeniería de características pero no modela la dinámica emocional. Por otro lado, el DEKT [3], que pertenece a la familia del *Deep Knowledge Tracing (DKT)*, utiliza una red neuronal dual para incorporar cuatro emociones (concentración, aburrimiento, confusión y frustración). Tal como detallan en su artículo, DEKT no es una extensión del BKT sino un modelo de caja negra que proporciona una baja interpretabilidad, el cual ofrece un alto rendimiento predictivo a cambio de necesitar miles de estudiantes para su entrenamiento, siendo inviable cuando los datos emocionales son escasos. Frente a estos, los modelos RE-BKT constituyen extensiones del *BKT* que mantienen la estructura bayesiana y la capacidad de operar con conjuntos de datos reducidos (25 estudiantes / 6 interacciones). La versión del Enfoque 1 expande la matriz de observación de 2 a 8 dimensiones para reflejar cuatro estados afectivos, lo que aumenta la carga computacional pero conserva una alta interpretabilidad paramétrica. La versión del Enfoque 2 modula la predicción mediante una función logística basada únicamente en la concentración, lo que simplifica el modelo pero obliga a un clustering previo para definir perfiles de sensibilidad emocional. Finalmente, la versión del Enfoque 3 combina ambos mecanismos (matriz expandida más modulación logística), logrando una interpretabilidad media-alta pero con un alto acoplamiento que puede magnificar los errores si alguno de los subsistemas no está bien calibrado.

Modelo	Variable contextual	Arquitectura base	Datos emocionales	Requerimiento de datos	Interpretabilidad paramétrica	Limitación principal
BKT [30]	Ninguna	Modelo BKT (Probabilístico)	Ninguno	25 estud. / 6 interacc.	Alta (parámetros clásicos)	Ignora por completo información afectiva o contextual no modelada explícitamente.
FAST [14]	Múltiples características (indicadores binarios de presencia/ausencia, dificultad del ítem, habilidad del estudiante, número de prácticas previas)	Regresión logística en emisiones	Ninguno	Bajo si hay buena ing. de características	Media-alta (coeficientes interpretables)	Requiere ingeniería de características; no modela dinámica emocional.
DEKT [31]	Emociones	Red neuronal dual (conocimiento + emoción)	Concentración, aburrimiento, confusión, frustración	Alto (miles de estudiantes)	Baja (modelo de caja negra)	Depende de grandes volúmenes; inviable con datos emocionales escasos.
RE-BKT - Enfoque 1	Dato Emocional	Expande matriz de observación	Solo concentración (otras emociones solo para clustering)	25 estud. / 6 interacc.	Alta (parámetros BKT + interpretación directa de la matriz expandida)	Complejidad al aumentar la matriz de estados (de 2 a 8 dimensiones), lo que incrementa la carga computacional.
RE-BKT - Enfoque 2		Modula predicción vía función logística		25 estud. / 6 interacc. (efectivo si hay alta correlación)	Alta (parámetros BKT + peso emocional)	Solo aprovecha la concentración; requiere clustering previo para definir perfiles de sensibilidad emocional.
RE-BKT - Enfoque 3		Matriz expandida + modulación Logística		25 estud. / 6 interacc.	Media-Alta (combina la interpretabilidad de ambos módulos)	Alto acoplamiento de componentes, sin embargo puede magnificar errores si uno no está bien calibrado.

Tabla 4.13. Comparación de modelos de seguimiento del conocimiento.

Nota: la comparación cuantitativa directa no es posible porque cada trabajo utiliza datasets, protocolos de evaluación y métricas distintas. Esta tabla ofrece una visión cualitativa de las mejoras, no una comparación en condiciones idénticas.

La Tabla 4.13 evidencia que los modelos RE-BKT se posicionan como una alternativa equilibrada y contextualmente viable frente a los enfoques de caja negra como DEKT. Al mantener la estructura bayesiana del BKT clásico, ofrecen un alto grado de interpretabilidad paramétrica y no requieren volúmenes masivos de datos para su entrenamiento, operando eficazmente con conjuntos reducidos.

Capítulo V

Conclusiones

Este proyecto de grado tuvo como objetivo principal extender el modelo BKT para incorporar las emociones en el proceso de inferencia del dominio de habilidades, proponiendo y evaluando tres arquitecturas diferenciadas: Enfoque 1: dato emocional en el nodo de observación, Enfoque 2: dato emocional en la predicción y Enfoque 3: Híbrido. Los hallazgos obtenidos permiten formular un conjunto de conclusiones ordenadas desde los aspectos más generales hasta las implicaciones más específicas del trabajo. Asimismo, se presentan los aportes, recomendaciones, limitaciones y trabajos futuros.

5.1 Conclusiones

Derivado del objetivo planteado, en primer lugar, se concluye que la forma en la que se incorpora el estado emocional del estudiante en el modelo BKT determina si aporta beneficio o degradación al modelo. La arquitectura del Enfoque 1, que integra la concentración directamente en el espacio discreto de observaciones del HMM expandiendo la matriz B , produjo mejoras en la especificidad respecto al modelo base, pero degradó sistemáticamente el AUC en todos los escenarios evaluados (ΔAUC entre $-0,0715$ y $-0,0884$ en ASSISTments 2012, y entre $-0,0203$ y $-0,0492$ en ASSISTments 2017 Global). Esto revela que forzar la emoción como un estado observable adicional en el proceso de entrenamiento introduce rigidez estructural que el algoritmo *Baum-Welch* no puede compensar con datos limitados (subconjunto Controlado de cada dataset).

Por el contrario, la arquitectura del Enfoque 2, que preserva intacta la estructura de entrenamiento del BKT y aplica una modulación logística de la concentración continua en la fase de predicción, produjo mejoras estadísticamente significativas en la capacidad discriminativa en siete de los ocho escenarios evaluados con la prueba de DeLong ($p < 0,0001$), con incrementos de AUC que oscilan entre $+0,0071$ en ASSISTments 2012 Global y $+0,0953$ en ASSISTments 2017 Controlado. Lo que sugiere que la concentración actúa como un modulador de la confianza en la predicción, no como un estado oculto

adicional, ya que al ajustar la probabilidad de acierto a posteriori mediante la función logística se corrige el sesgo del modelo BKT clásico hacia la clase mayoritaria (en ASSISTments 2012, el modelo tiende a sobre predecir respuestas correctas debido al desbalance hacia los aciertos; en ASSISTments 2017, tiende a sub predecir los aciertos porque predominan las respuestas incorrectas), sin alterar su estructura bayesiana fundamental. En cuanto a la arquitectura Híbrida, que combina ambos enfoques, presentó un rendimiento intermedio, en ASSISTments 2012 donde la correlación con el dato emocional era débil el modelo degradó el AUC en todos los escenarios, presentando diferencias negativas significativas; sin embargo, en ASSISTments 2017 mejoró el AUC pero no alcanzó la eficiencia de la versión del Enfoque 2, y sus mejoras en el subconjunto Controlado no alcanzaron significancia estadística. En consecuencia, la recomendación derivada de esta investigación es que para incorporar la emoción al BKT con beneficio predictivo garantizado, debe hacerse en la fase de predicción como modulación externa, no en la fase de entrenamiento como observación discreta adicional.

En segundo lugar, uno de los supuestos a verificar era si la ventaja del modelo emocional se mantiene cuando los datos son escasos, como se reporta en la Sección 4.4.2, en el subconjunto Controlado de ASSISTments 2012 (restringido a 25 estudiantes con las primeras 6 interacciones por habilidad), las mejoras de la arquitectura del Enfoque 2 respecto al BKT base no alcanzaron significancia estadística ($\Delta AUC = +0,0087$ en clúster 0 con $p = 0,4145$ y $-0,0065$ en clúster 1 con $p = 0,3259$). Esto indica que, cuando la correlación entre emoción y rendimiento es débil (como en ASSISTments 2012, donde los coeficientes punto-biserial no superan 0,12), la señal emocional es insuficiente para marcar una diferencia estadísticamente robusta con tan pocas observaciones. Sin embargo, en el subconjunto Controlado de ASSISTments 2017, donde los coeficientes de correlación emoción-rendimiento son notablemente más fuertes (concentración: 0,35), la arquitectura del Enfoque 2 logró las mejoras más pronunciadas de toda la investigación incluso con datos limitados ($\Delta AUC = +0,0953$ en clúster 0 y $+0,0609$ en clúster 1, ambas con $p < 0,0001$). Este hallazgo demuestra que el RE-BKT mantiene un rendimiento competitivo incluso con muestras pequeñas únicamente cuando la correlación emocional es suficiente fuerte ($r \geq 0,30$). El dataset ASSISTments 2017 se caracteriza precisamente por presentar una correlación más intensa entre la concentración y el rendimiento académico, lo que refleja una

mayor sensibilidad emocional de los estudiantes en ese contexto. En contraste, los modelos de redes neuronales como los de [3][11] colapsan o no convergen en ese rango de datos, pues requieren volúmenes masivos de interacciones para evitar el sobreajuste.

En tercer lugar, se concluye que la concentración es el dato emocional con mayor valor predictivo para el rendimiento académico en ambos conjuntos de datos analizados, presentando los coeficientes de correlación punto-biserial más elevados y consistentes. Esta observación es coherente con la Teoría del Flujo de Csikszentmihalyi [22] [23], que describe la concentración total en la tarea como la característica definitoria del estado óptimo de aprendizaje. Las emociones negativas (frustración, confusión y aburrimiento) mostraron correlaciones más débiles o inconsistentes entre conjuntos de datos, lo que justificó la decisión metodológica de reservarlas exclusivamente para la fase de segmentación de perfiles estudiantiles y no incorporarlas directamente como variables predictoras en la función de emisión del modelo.

En cuarto lugar, se concluye que la estrategia de personalización por perfil de sensibilidad emocional, implementada mediante la segmentación *K-means* sobre un vector de características por estudiante, constituye un mecanismo efectivo para capturar la heterogeneidad en la relación entre emoción y aprendizaje. La evidencia experimental demostró que los parámetros óptimos del modelo difieren de forma sustancial entre los dos clústeres identificados, lo que valida la premisa de que asumir parámetros homogéneos para toda la población estudiantil introduce un error de especificación sistemático en el modelo.

Finalmente, se concluye que el BKT clásico exhibe en varios escenarios una especificidad extremadamente baja (hasta 0,0855), lo que significa que el modelo predice dominio de la habilidad incluso cuando el estudiante falla varias veces. De esta manera, el modelo válida incorrectamente el aprendizaje de habilidades no dominadas. El RE-BKT con incorporación del dato emocional en la predicción corrige este sesgo elevando la especificidad a rangos entre 0,60 y 0,81 sin afectar de forma abrupta la sensibilidad, gracias a la optimización de umbrales por clúster y la modulación logística de la concentración. En cambio, el RE-BKT con incorporación del dato emocional en el nodo de observación y el RE-BKT Híbrido no logran esta corrección de manera consistente: el primero degrada el AUC en todos los escenarios, mientras que el segundo modelo sólo mejora cuando la correlación emocional es fuerte (ASSISTments 2017), pero sin alcanzar los niveles de

equilibrio entre sensibilidad y especificidad que proporciona la versión externa. Por lo tanto, la corrección del sesgo del BKT solo es efectiva cuando la emoción se incorpora externamente mediante regresión logística, y no cuando se fuerza dentro del espacio de observación del modelo.

5.2 Aportes

Las principales contribuciones de este proyecto de grado son las siguientes:

- Se desarrollaron y se implementaron tres arquitecturas de extensión emocional del BKT, las cuales fueron: El RE-BKT con incorporación del dato emocional en el nodo de observación, el RE-BKT con incorporación del dato emocional en la predicción, y el RE-BKT Híbrido. Específicamente, se contribuyó extendiendo las ecuaciones bayesianas y las ecuaciones de re-estimación de la matriz ampliada, y proponiendo una función de predicción logística.
- Se diseñó e implementó una arquitectura de entrenamiento diferenciado por perfil emocional, que ajusta los parámetros del modelo de forma independiente para cada clúster de estudiantes.
- Se realizó un análisis estadístico de significancia mediante la prueba de DeLong sobre ocho escenarios de evaluación ($4 \text{ subconjuntos} \times 2 \text{ datasets}$), proporcionando evidencia cuantitativa rigurosa sobre cuándo y bajo qué condiciones la incorporación emocional genera mejoras genuinas y no atribuibles al azar.

5.3 Limitaciones

Durante la realización de este proyecto de grado se presentaron ciertas limitaciones que deben tenerse en cuenta al momento de interpretar los resultados obtenidos. En primer lugar, la escasez de conjuntos de datos con información emocional fue una limitación significativa, ya que se realizó una búsqueda extensa y solo se encontraron dos datasets públicos que incorporan variables emocionales de los estudiantes. La gran mayoría de los repositorios disponibles en minería de datos educativos no consideran esta dimensión, lo que restringe las posibilidades de validación y comparación del modelo propuesto en contextos adicionales.

En segundo lugar, los datasets públicos conseguidos que contienen información emocional, específicamente los generados por la plataforma ASSISTments, no proporcionan las señales emocionales sin procesar, sino que ya incluyen valores promediados estimados por el propio sistema. Lo que implica que sólo es posible experimentar sobre estos promedios sin aplicar transformaciones, limitando el control sobre los datos emocionales.

Aunado a esto, la propuesta desarrollada se restringe al uso de la concentración como único dato emocional en la función de predicción del modelo RE-BKT. Esta decisión se fundamentó en el análisis de correlación punto-biserial, que mostró que la concentración era el único dato emocional con una correlación positiva y consistentemente significativa con el rendimiento académico en todos los conjuntos de datos.

5.4 Recomendaciones

A partir de los hallazgos y limitaciones de esta investigación, se formulan las siguientes recomendaciones:

- Se recomienda adoptar el Enfoque 2: dato emocional en la predicción, como configuración predeterminada del RE-BKT en nuevas aplicaciones, dada su superioridad consistente y estadísticamente validada frente a las variantes de los Enfoques 1 y 3. La incorporación de la emoción en la fase de predicción, mediante la función logística con parámetros ω y η optimizados por *Optuna*, ofrece el mejor equilibrio entre ganancia predictiva y complejidad computacional.
- Se recomienda explorar esquemas de segmentación específicos para cada conjunto de datos con información emocional, ya que la formación de los perfiles de sensibilidad emocional no es necesariamente estática y debe personalizarse según la naturaleza de los datos.
- Se recomienda que los diseñadores de sistemas tutores inteligentes prioricen la medición del nivel de concentración del estudiante, dado que esta variable demostró el mayor valor predictivo sobre el rendimiento académico, por encima de otras emociones como la frustración, confusión o el aburrimiento.

5.5 Trabajo Futuro

A partir de los resultados obtenidos, se identifican las siguientes líneas de trabajo futuro:

En primer lugar, la línea de investigación más inmediata y directamente derivada de esta tesis es el desarrollo del modelo FE-BKT (Fluctuación Emocional - *Bayesian Knowledge Tracing*). Esta extensión debería capturar la dinámica temporal de la transición entre estados emocionales descrita por [21], modelando explícitamente cómo las secuencias de estados emocionales-cognitivos (por ejemplo, transiciones de concentración a frustración o de aburrimiento a compromiso) impactan en la trayectoria de adquisición del conocimiento.

Segundo, un trabajo futuro de interés es evaluar el RE-BKT con incorporación del dato emocional en la predicción en conjuntos de datos provenientes de áreas distintas a las matemáticas, como el aprendizaje de lenguas, ciencias naturales o programación. La relación entre emociones y rendimiento podría diferir según la naturaleza de la habilidad, por lo que sería valioso determinar si la concentración sigue siendo el predictor dominante o si otras emociones (como la frustración ante errores sintácticos en programación) adquieren mayor relevancia predictiva. Este análisis permitiría delimitar el dominio de validez del RE-BKT y eventualmente ajustar su función de modulación emocional al contexto específico.

En términos de modelado computacional, se plantea como línea futura la exploración de arquitecturas híbridas que combinen el RE-BKT con incorporación del dato emocional en la predicción, con redes neuronales recurrentes (*LSTM* o mecanismos de atención tipo *Transformer*) para modelar dependencias temporales de orden superior en la secuencia de interacciones, manteniendo al mismo tiempo la interpretabilidad bayesiana del núcleo del modelo.

Para sistemas tutores, una línea de integración natural es combinar el RE-BKT con aprendizaje por refuerzo (*RL*) para acelerar la estimación del conocimiento de los estudiantes [34] [35] o para adaptar intervenciones según el conocimiento y el estado emocional del estudiante [36]. En este sentido, el RE-BKT podría actuar como un módulo de diagnóstico cognitivo-emocional dentro de un sistema tutor basado en *RL*, proporcionando al agente tanto la estimación del dominio de la habilidad como información del impacto emocional.

Referencias

- [1] A. T. Corbett and J. R. Anderson, “Knowledge tracing: Modeling the acquisition of procedural knowledge,” *User Model. User-Adapt. Interact.*, vol. 4, pp. 253–278, 1994.
- [2] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon, “Individualized bayesian knowledge tracing models,” in *Proc. 16th Int. Conf. Artif. Intell. Educ.*, Memphis, TN, USA, 2013, pp. 171–180.
- [3] S. Wang et al., “Dual-State Personalized Knowledge Tracing with Emotional Incorporation,” *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 5, pp. 2440–2455, 2025.
- [4] S. Slater and R. S. Baker, “Degree of Error in Bayesian Knowledge Tracing Estimates From Differences in Sample Sizes,” *Behaviormetrika*, vol. 45, no. 3, pp. 707–727, 2018.
- [5] C. E. Dowling and C. Hockemeyer, “Automata for the assessment of knowledge,” *IEEE Trans. Knowl. Data Eng.*, vol. 13, no. 3, pp. 451–461, May 2001.
- [6] Z. A. Pardos et al., “Adapting bayesian knowledge tracing to a massive open online course in edx,” in *Proc. 6th Int. Conf. Educ. Data Min.*, Memphis, TN, USA, 2013, pp. 308–309
- [7] F. Wang et al., “NeuralCD: A general framework for cognitive diagnosis,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 6, pp. 6436–6449, Jan. 2022.
- [8] S. Spaulding, G. Gordon, and C. Breazeal, “Affect-aware student models for robot tutors,” in *Proc. 15th Int. Conf. Auton. Agents Multiagent Syst. (AAMAS)*, Singapore, May 9–13, 2016, pp. 864–872.
- [9] Q. Liu et al., “A Survey of Knowledge Tracing: Models, Variants, and Applications,” *IEEE Transactions on Learning Technologies*, vol. 17, 2024, (Survey Paper)
- [10] Y. Xu, K. Chang, Y. Yuan, and J. Mostow, “EEG helps knowledge tracing!,” in *Workshop Util. EEG Input Intell. Tutoring Syst. (ITS)*, Honolulu, HI, USA, 2014, pp. 43–50

- [11] S. Schultz and I. Arroyo, “Tracing knowledge and engagement in parallel in an intelligent tutoring system,” in *Proc. 7th Int. Conf. Educ. Data Min.*, London, UK, 2014, pp. 242–247.
- [12] Y. Wang and N. T. Heffernan, “Leveraging first response time into the knowledge tracing model,” in *Proc. 5th Int. Conf. Educ. Data Min.*, Chania, Greece, 2012, pp. 168–171.
- [13] J. E. Beck et al., “Does help help? Introducing the Bayesian Evaluation and Assessment Methodology,” in *Proc. 9th Int. Conf. Intell. Tutoring Syst.*, Montreal, QC, Canada, 2008, pp. 383–394.
- [14] J. P. González-Brenes, Y. Huang, and P. Brusilovsky, “General features in knowledge tracing to model multiple subskills, temporal item response theory, and expert knowledge”, in *Proc. 7th Int. Conf. Educ. Data Min.*, London, UK, 2014, pp. 84–91.
- [15] L. Salazar Gómez, A. A. Vázquez Sánchez, y R. Cañizares González, “Estimar conocimiento latente en grandes volúmenes de datos utilizando el algoritmo Bayesian Knowledge Tracing”, *Revista Cubana de Ciencias Informáticas*, vol. 15, especial UCIENCIA I, sept. 2021.
- [16] D. Shchepakina, S. Sankaranarayanan, y D. Zimmaro, “Parametric constraints for Bayesian Knowledge Tracing from first principles,” in *Proceedings of the 17th International Conference on Educational Data Mining*, B. Paaßen y C. D. Epp, Eds., Atlanta, GA, USA, Jul. 2024, pp. 18–29. International Educational Data Mining Society. CC BY-NC-ND 4.0 International, 2024
- [17] R. Pekrun, T. Goetz, W. Titz, y R. P. Perry, “Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research,” *Educational Psychologist*, vol. 37, n.º 2, pp. 91-106, 2002.
- [18] G. H. Bower, “Mood and memory,” *American Psychologist*, vol. 36, n.º 2, pp. 129-148, 1981.
- [19] R. Pekrun, “Control-Value Theory: From Achievement Emotion to a General Theory of Human Emotions,” *Educational Psychology Review*, vol. 36, n.º 83, 2024.

- [20] R. Pekrun, “The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice,” *Educational Psychology Review*, vol. 18, n.º 4, pp. 315-341, 2006.
- [21] S. D'Mello y A. Graesser, “Dynamics of affective states during complex learning,” *Learning and Instruction*, vol. 22, n.º 2, pp. 145-157, 2012.
- [22] L. Li-Fen Liao, “A Flow Theory Perspective on Learner Motivation and Behavior in Distance Education,” *Distance Education*, vol. 27, no. 1, pp. 45–62, 2006.
- [23] W. O. dos Santos, I. I. Bittencourt, D. Dermeval, S. Isotani, L. B. Marqués, y I. F. Silveira, “Flow Theory to Promote Learning in Educational Systems: Is it Really Relevant?” *Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação - RBIE)*, vol. 26, no. 2, pp. 29–59, 2018.
- [24] Z. A. Pardos, R. S. J. D. Baker, M. O. C. Z. San Pedro, S. M. Gowda, y S. M. Gowda, “Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes,” en *Proc. 3rd Int. Conf. Learn. Analytics and Knowledge (LAK)*, Leuven, Bélgica, abr. 2013, pp. 117–124.
- [25] ASSISTments, “ASSISTments Data Mining Challenge 2017 Dataset,” ASSISTments Data Mining. [En línea]. Disponible en: <https://sites.google.com/view/assistmentsdatamining/dataset>. [Accedido: 16-dic-2025].
- [26] Z. A. Pardos y N. T. Heffernan, “Navigating the parameter space of Bayesian Knowledge Tracing models: Visualizations of the convergence of the Expectation Maximization algorithm”, en *Proceedings of the 3rd International Conference on Educational Data Mining (EDM 2010)*, Pittsburgh, PA, 2010, pp. 161-170.
- [27] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006, cap. 13, pp. 605-634.
- [28] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, “Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach,” *Biometrics*, vol. 44, no. 3, pp. 837–845, Sep. 1988.

- [29] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining (KDD '19)*, Anchorage, AK, USA, Aug. 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.
- [30] M. Baktashmotlagh, Y. Wang, and K. R. Koedinger, “pyBKT: An accessible Python library for Bayesian Knowledge Tracing,” in *Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021)*, I. Roll, D. McNamara, S. Sosnovsky, and R. Luckin, Eds., International Educational Data Mining Society, 2021, pp. 518–524.
- [31] S. Trivedi, Z. A. Pardos, and N. T. Heffernan, “clustering Students to Generate an Ensemble to Improve Standard Test Score Predictions,” in *Artificial Intelligence in Education*, G. Biswas, S. Bull, J. Kay, and A. Mitrovic, Eds., Berlin, Heidelberg: Springer, 2011, pp. 377–384. DOI: 10.1007/978-3-642-21869-9_49.
- [32] D. Zambrano and J. Pérez, “Diseño de un Modelo para Predecir el Rendimiento de la Memoria de Trabajo en Estudiantes Universitarios según los niveles de Atención y Estrés Mental,” 2025. doi: 10.13140/RG.2.2.28591.65442.
- [33] C. H. Zelada, “Modelo Lineal Generalizado (GLM),” *RPubs*, 2017. [Online]. Available: <https://rpubs.com/chzelada/275494>. [Accessed: Apr. 16, 2026].
- [34] J. Pérez, E. Dapena, J. Aguilar, and G. Carrillo, “Reinforcement Learning for Estimating Student Proficiency in Math Word Problems,” in *2022 XVII Latin American Conference on Learning Technologies (LACLO)*, Armenia, Colombia, 2022, pp. 01–06.
- [35] J. Pérez, J. Aguilar, and E. Dapena, “Q-learning based algorithm for learning children’s difficulties in multiplication tables,” *Revista Ciencia e Ingeniería*, vol. 46, no. 3, pp. 343–352, 2025.
- [36] J. Pérez, E. Dapena, and J. Aguilar, “Emotions as implicit feedback for adapting difficulty in tutoring systems based on reinforcement learning,” *Education and Information Technologies*, 2024. doi: 10.1007/s10639-024-12699-8.

[37] O. Pedroza, “ExtensionBKT”, GitHub. [Online]. Available: <https://github.com/orianapedroza/ExtensionBKT.git>. [Accessed: Abr. 12, 2026].

Apéndice

Apéndice A: Análisis Exploratorio de Datos

Este apéndice complementa la sección 4.2 del Capítulo IV, presentando de forma exhaustiva el preprocesamiento, el análisis de correlaciones, el proceso de clustering (*K-means*), la validación del número de clústeres mediante codo y silueta, la visualización por componentes principales y la caracterización estadística completa de los perfiles estudiantiles identificados en los dos conjuntos de datos. Por ende el *EDA* permitió comprender la estructura y distribución de las variables en los dos conjuntos de datos seleccionados: ASSISTments 2012-13 (Dataset 1) y ASSISTments Challenge 2017 (Dataset 2). El enfoque principal fue la identificación de patrones entre el desempeño académico y los estados emocionales-cognitivos previstos, tales como concentración, frustración, confusión y aburrimiento.

A.1 Preprocesamiento de los conjuntos de datos

Para garantizar la calidad de la inferencia en los modelos de seguimiento del conocimiento, se aplicó un protocolo de limpieza y preprocesamiento para cada dataset, como se observa a continuación.

A.1.1 Preprocesamiento del Dataset 1 (ASSISTments 2012-2013)

En primera instancia, se procedió a la remoción de dimensiones irrelevantes para los objetivos de esta investigación, tales como identificadores de clase, escuela, profesor y parámetros contextuales de la interfaz (*tutor_mode*, *answer_id*, entre otros).

```
irrelevant_columns = ['problem_id', 'assignment_id', 'problem_type', 'original', 'bottom_hint', 'answer_id', 'answer_text',
                      'hint_count', 'actions', 'ms_first_response', 'tutor_mode', 'student_class_id', 'position', 'type',
                      'base_sequence_id', 'overlap_time', 'teacher_id', 'school_id', 'problemlogid', 'first_action',
                      'assistment_id', 'sequence_id', 'template_id',]
```

Seguidamente se estandarizó la columna *correct* mediante una conversión de tipos para asegurar que los valores se restringen estrictamente al rango binario {0,1}. También se estandarizó el tipo de dato de las columnas *start_time* y *end_time* garantizando la

secuencia cronológica de las interacciones. En cuanto a los registros nulos, se obtuvo la columna `skill` con 3.493.190 valores nulos y en la columna `skill_id` con 3.411.457, afectando así 9.660 estudiantes. Como se observa en el Fig. 4.2, la magnitud de estos vacíos impedía una asociación válida entre la tarea y la competencia evaluada.

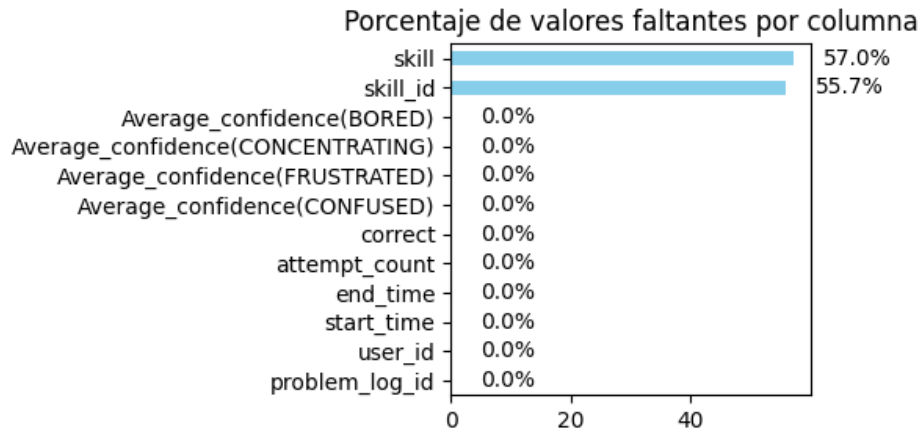


Fig. A.1. Distribución porcentual de valores nulos por variable (Dataset 1).

Tras la depuración de estos registros (3.493.190 interacciones eliminadas), el conjunto de datos resultante conservó 2.630.080 interacciones correspondientes a 28.834 estudiantes únicos y 197 habilidades válidas. Seguido de esto, se definieron las variables fundamentales para la implementación: identificador de usuario (`user_id`), la habilidad (`skill`), corrección de la respuesta (`correct`) y las cuatro variables emocionales continuas (`Average_confidence(CONCENTRATING)`, `Average_confidence(FRUSTRATED)`, `Average_confidence(CONFUSED)` y `Average_confidence(BORED)`).

A.1.2 Preprocesamiento del Dataset 2 (ASSISTments 2017)

Para el segundo conjunto de datos, se aplicó un proceso análogo de depuración de columnas no esenciales para el entrenamiento del modelo de conocimiento:

```

irrelevant_columns = ['SY ASSISTments Usage', 'AveCarelessness', 'AveResOfftask', 'AveResGaming',
                      'problemId', 'assignmentId', 'assistentId', 'original', 'hint', 'hintTotal', 'scaffold',
                      'bottomHint', 'frIsHelpRequest', 'frPast5HelpRequest', 'frPast8HelpRequest', 'stlHintUsed',
                      'past8BottomOut', 'totalFrPercentPastWrong', 'totalFrPastWrongCount', 'frPast5WrongCount',
                      'frPast8WrongCount', 'totalFrTimeOnSkill', 'frWorkingInSchool', 'totalFrAttempted', 'responseIsFillIn',
                      'responseIsChosen', 'endsWithScaffolding', 'endsWithAutoScaffolding', 'frTimeTakenOnScaffolding',
                      'frTotalSkillOpportunitiesScaffolding', 'totalFrSkillOpportunitiesByScaffolding',
                      'frIsHelpRequestScaffolding', 'timeGreater5Secprev2wrong', 'sumRight', 'helpAccessUnder2Sec',
                      'timeGreater10SecAndNextActionRight', 'consecutiveErrorsInRow', 'sumTime3SDWhen3RowRight',
                      'sumTimePerSkill', 'totalTimeByPercentCorrectForSkill', 'timeOver80', 'manywrong',
                      'confidence(BORED)', 'confidence(CONCENTRATING)', 'confidence(CONFUSED)', 'confidence(FRUSTRATED)',
                      'confidence(OFF TASK)', 'confidence(GAMING)', 'RES_OFFTASK', 'RES_GAMING', 'Prev5count',
                      'InferredGender', 'Selective', 'isSTEM', 'Enrolled', 'MiddleSchoolId',
                      'AveKnow', 'AveCorrect', 'NumActions', 'action_num', 'problemType', 'timeTaken', 'hintCount',
                      'timeSinceSkill', 'totalFrSkillOpportunities', 'MCAS', 'Ln-1', 'Ln', 'AveResBored',
                      'AveResEngcon', 'AveResConf', 'AveResFrustr']

```

A diferencia del primer dataset, la inspección de la variable `correct` confirmó el cumplimiento del formato binario esperado, y la normalización temporal no reveló discrepancias en la secuencia de eventos. De igual forma, el análisis de valores faltantes indicó que este dataset posee una integridad total, careciendo de registros nulos en sus dimensiones. Para esta implementación, se procesaron los datos completos y se seleccionaron las variables: identificador del estudiante (`studentId`), la habilidad (`skill`), corrección de la respuesta (`correct`) y las cuatro variables emocionales continuas (`RES_CONCENTRATING`, `RES_FRUSTRATED`, `RES_CONFUSED` y `RES_BORED`).

A.2 Análisis de Correlación

Con el objetivo de explorar las relaciones lineales entre las emociones reportadas por los estudiantes y su rendimiento académico (respuesta correcta), se realizó un análisis de correlación para cada conjunto de datos por separado. Dado que la variable de rendimiento es dicotómica (correcto/incorrecto) y las emociones son variables continuas, se empleó el coeficiente de correlación point-biserial, que es equivalente al coeficiente de Pearson en este contexto. Los resultados se presentan en matrices de correlación y gráficos de barras.

A.2.1 Dataset 1: ASSISTments 2012-2013

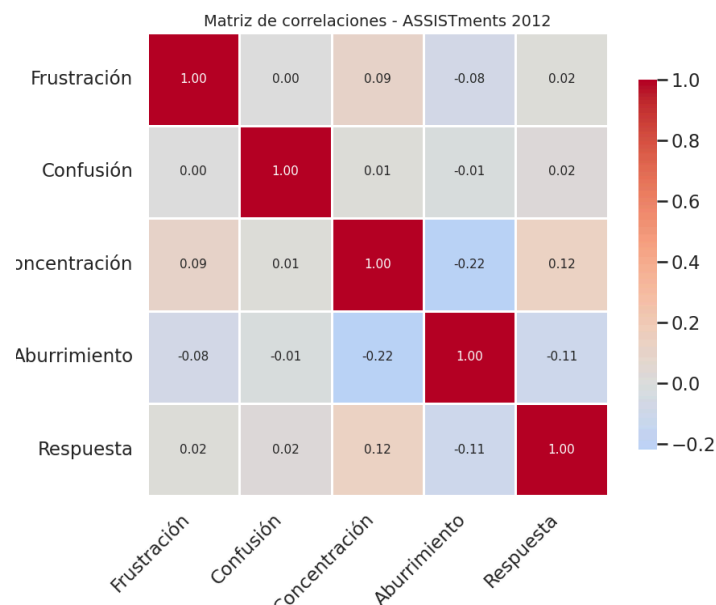


Fig. A.2. Coeficientes de Correlación Punto-Biserial ASSISTments 2012.

En el conjunto de datos ASSISTments 2012-2013, la matriz de correlación (Fig. A.2) muestra que el dato emocional con mayor asociación positiva con la respuesta correcta es la concentración, con un coeficiente de 0,12. Esto indica una relación débil pero positiva, sugiriendo que los estudiantes que reportan mayores niveles de concentración tienden a obtener mejores resultados. En cambio, el aburrimiento muestra una correlación negativa de $-0,11$ confirmando que el desinterés se vinculó con una menor probabilidad de acierto. Por otro lado, tanto la frustración como la confusión presentan coeficientes prácticamente nulos (0,02), lo que indica la ausencia de una relación lineal apreciable entre estos estados y el éxito inmediato en este dataset específico. Adicionalmente, se halló una correlación negativa moderada entre concentración y aburrimiento ($r = -0,22$), lo cual resultó coherente con la naturaleza excluyente de ambos estados. El resto de las intercorrelaciones son cercanas a cero, lo que sugiere que las cuatro variables emocionales miden dimensiones relativamente independientes entre sí.

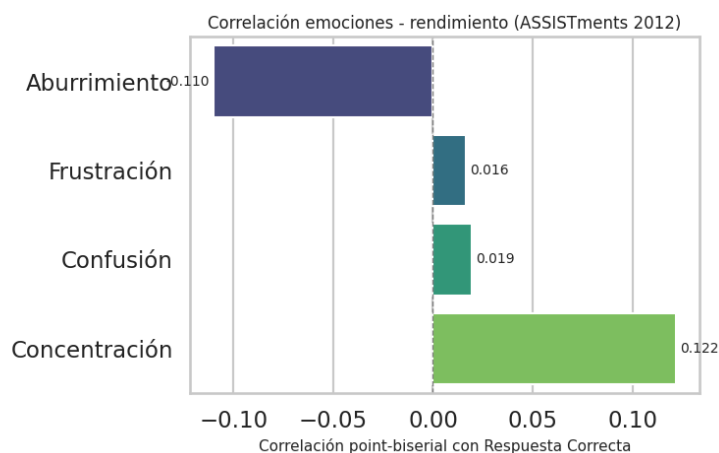


Fig. A.3. Gráfico de barras de correlaciones ASSISTments 2012.

En cuanto a la Figura A.3 se muestra de forma más sintética los coeficientes de correlación punto-biserial entre cada dato emocional y la respuesta correcta en este dataset, se observa que la concentración es el dato con la correlación positiva más alta (0,12), seguida de valores prácticamente nulos para frustración y confusión (0,02 cada una). El aburrimiento aparece con una correlación negativa de $-0,11$, destacándose como el único estado afectivo con una asociación claramente inversa al rendimiento. La disposición de las barras permite visualizar que de los cuatro datos emocionales considerados, únicamente la concentración y el aburrimiento muestran una relación apreciable (aunque débil) con la probabilidad de acierto en este conjunto de datos, mientras que la frustración y la confusión no presentan una influencia lineal significativa sobre el éxito inmediato del estudiante.

A.2.2 Dataset 2: ASSISTments Challenge 2017

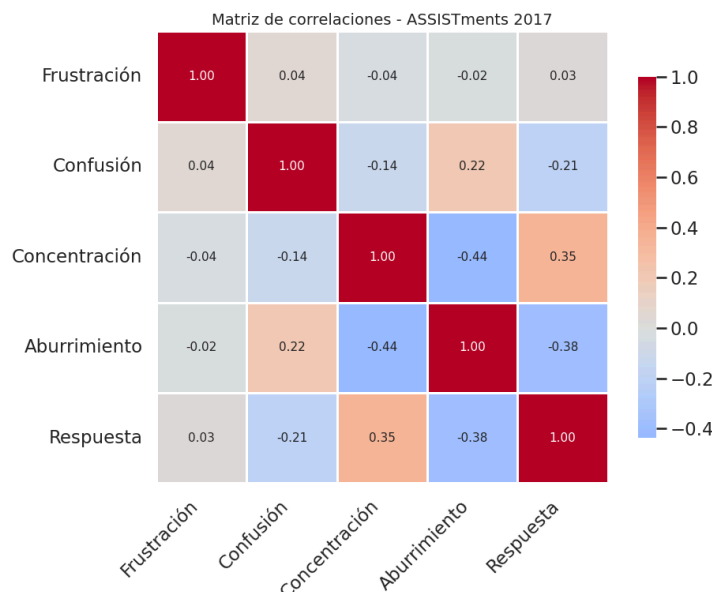


Fig. A.4. Coeficientes de Correlación Punto-Biserial ASSISTments 2017.

El análisis del conjunto ASSISTments Challenge 2017 revela patrones de correlación notablemente diferentes al dataset anterior, ver Figura A.4. La concentración vuelve a mostrar una correlación positiva con el rendimiento, pero esta vez alcanza un valor de 0,35, considerablemente más alto que el observado en 2012. De manera similar, el aburrimiento presenta una correlación negativa de $-0,38$, la más fuerte en magnitud entre todas las emociones analizadas. La confusión, que en el dataset anterior no mostraba relación con el rendimiento, ahora exhibe una correlación negativa de $-0,21$, indicando que los estudiantes confusos tienden a responder incorrectamente con mayor frecuencia. En cambio, la frustración mantiene una correlación prácticamente nula (0,03), consistente con los resultados del primer dataset. En cuanto a las relaciones entre emociones en ASSISTments 2017, se observa una correlación negativa fuerte entre concentración y aburrimiento ($r = -0,44$), una correlación positiva moderada entre confusión y aburrimiento ($r = 0,22$), y una correlación negativa débil entre confusión y concentración ($r = -0,14$). Estas intercorrelaciones internas justificaron la integración de estas variables emocionales para la personalización de los modelos RE-BKT, confirmando que la sensibilidad emocional es un factor determinante y diferenciado según el contexto del dataset analizado.

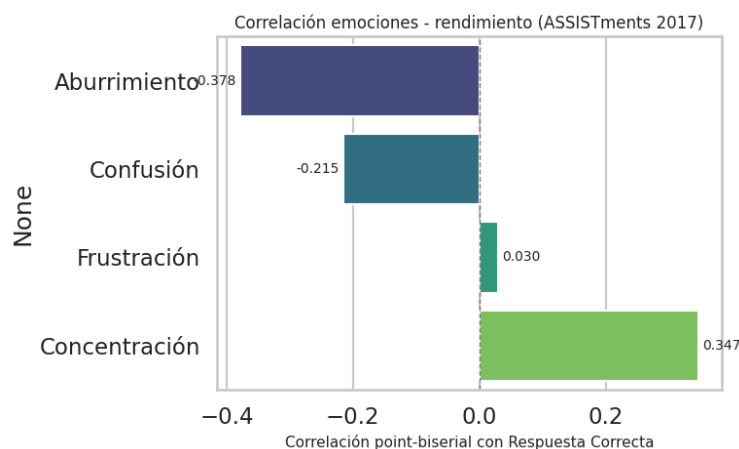


Fig. A.5. Gráfico de barras de correlaciones ASSISTments 2017.

Por su parte, la Figura A.5 presenta un gráfico de barras el cual confirma visualmente que la concentración es el dato emocional con la correlación positiva más alta (0,35), seguida de una correlación prácticamente nula para la frustración (0,03). La confusión aparece con una correlación negativa de $-0,21$, mientras que el aburrimiento muestra el valor más extremo con $-0,38$. La disposición de las barras permite apreciar rápidamente que, en este conjunto de datos, tanto la concentración como el aburrimiento tienen una influencia moderada sobre el rendimiento (positiva y negativa, respectivamente), mientras que la confusión comienza a mostrar un efecto negativo apreciable. La frustración, en cambio, carece de una relación lineal significativa con el éxito inmediato. Estos resultados contrastan con los del dataset ASSISTments 2012, donde las correlaciones eran más débiles, y evidencian que la fuerza de la relación emoción-rendimiento puede variar sustancialmente según el contexto educativo.

A.3 Definición de los perfiles estudiantiles

Con el fin de identificar patrones diferenciados de sensibilidad emocional, se realizó una caracterización de cada estudiante. Para ello, se generó un vector de 13 características que incluyó: la tasa de aciertos, la media y desviación estándar de los cuatro datos emocionales (concentración, frustración, confusión y aburrimiento) y correlación punto-biserial entre cada dato emocional y la respuesta correcta. La segmentación se ejecutó de forma independiente sobre dos subconjuntos de datos extraídos de ASSISTments 2012 y dos de ASSISTments 2017. Para cada dataset se construyó un conjunto “Global” (que incluye el historial completo

de estudiantes con al menos 6 interacciones) y un conjunto “Controlado” (restringido estrictamente a las primeras 6 interacciones de 25 estudiantes por habilidad). Esto permitió comparar el rendimiento y el comportamiento de los modelos tanto con una cantidad grande de datos como con una cantidad reducida.

A.3.1 Selección del algoritmo y optimización del número de clústeres

Para la segmentación de los perfiles se realizó la estandarización de las trece características mediante *StandardScaler* para asegurar que las diferencias en las escalas de las variables no sesgaran el cálculo y se llevó a cabo un estudio comparativo empleando diversos algoritmos de aprendizaje no supervisado. Se evaluaron modelos de agrupamiento basados en particiones, jerarquías, densidades y modelos probabilísticos, específicamente: *K-Means*, *clustering Jerárquico (Agglomerative)*, *DBSCAN*, *Gaussian Mixture Model (GMM)* y *HDBSCAN*. La validación de estos modelos se llevó a cabo mediante una proyección de los datos en dos dimensiones utilizando el *Análisis de Componentes Principales (PCA)*, permitiendo detallar la morfología de los grupos generados.

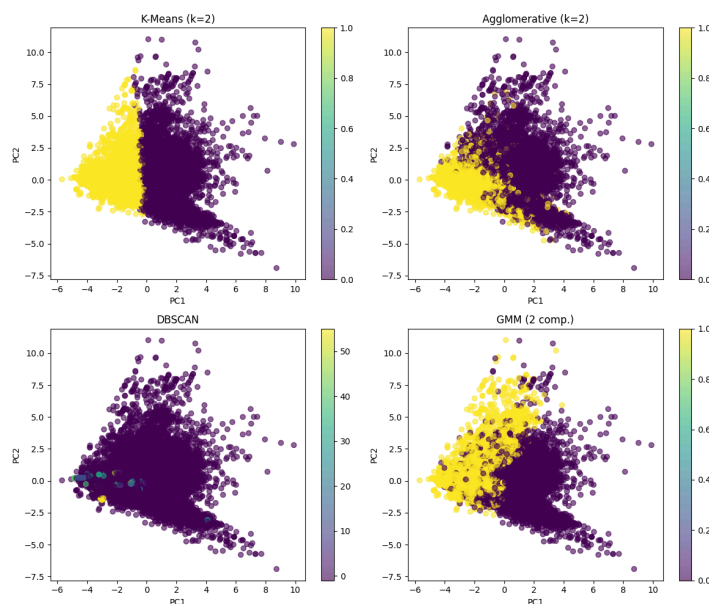


Fig. A.6. Evaluación de métodos de agrupamiento sobre los datos de emociones en ASSISTments 2012.

Como se observa en la comparativa de algoritmos Fig. A.6, el modelo *K-Means* proporcionó la separación más coherente y equilibrada de la población estudiantil tanto en el dataset 1 como en el dataset 2. Otros métodos, como *DBSCAN*, tendieron a clasificar la gran

mayoría de los puntos como un único grupo o ruido, mientras que *GMM* generó solapamientos excesivos que dificultan la interpretación pedagógica de los perfiles.

A.3.2 Metodología de agrupamiento: *K-Means*

El número óptimo de clústers se calculó mediante dos métricas complementarias: el *método del codo* (basado en inercia) y el *coeficiente de silueta*. Ambos procedimientos se aplicaron por separado para cada uno de los dos conjuntos de datos, considerando valores de k desde 2 hasta 10 similares a los usados en [31].

En el caso del primer conjunto de datos (ASSISTments 2012-2013), la Fig. A.7 muestra la evolución de la inercia y del coeficiente de silueta en función del número de clústeres. La inercia disminuye de manera pronunciada desde $k = 2$ hasta $k = 4$, y a partir de ese punto la reducción se vuelve gradual, lo que sugiere que la ganancia en compactación de los grupos se estabiliza a partir de cuatro clústeres. Sin embargo, el coeficiente de silueta, que mide la cohesión y separación de los agrupamientos, alcanza su valor máximo en $k = 2$ con una media de 0,172. Este comportamiento indica que la estructura más natural y bien definida del conjunto de datos es la partición en dos grupos, ya que una mayor fragmentación deteriora la calidad del agrupamiento.

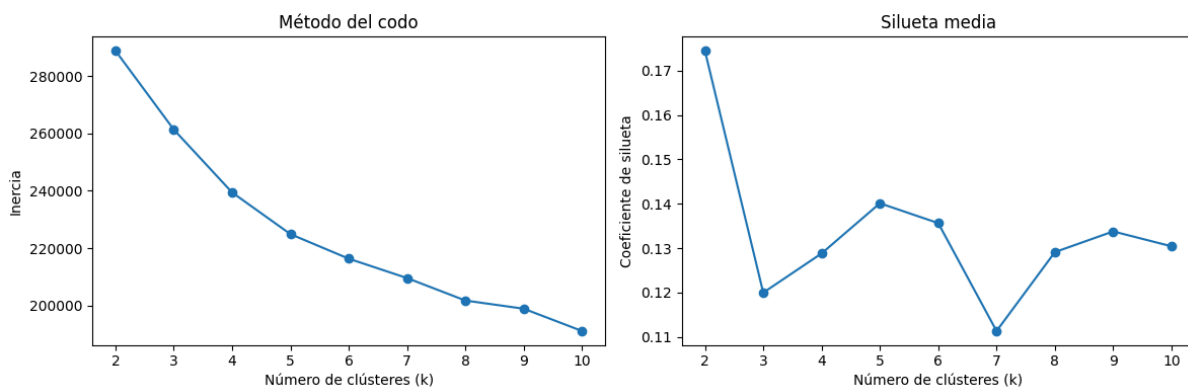


Fig. A.7. Análisis de codo (a) y silueta (b) para el agrupamiento del Dataset 1.

El segundo conjunto de datos (ASSISTments Challenge 2017) presentó un patrón similar, aunque con valores de silueta más elevados. Como se observa en la Fig. A.8, la inercia también desciende rápidamente hasta $k = 4$ aproximadamente y luego se estabiliza. El coeficiente de silueta muestra su punto máximo en $k = 2$, con un valor de 0,22, superando claramente a los obtenidos para el resto de los valores de k . Este máximo global confirma que

la partición en dos clústeres es la más adecuada para capturar la variabilidad subyacente en las respuestas emocionales-cognitivas de los estudiantes de esta muestra.

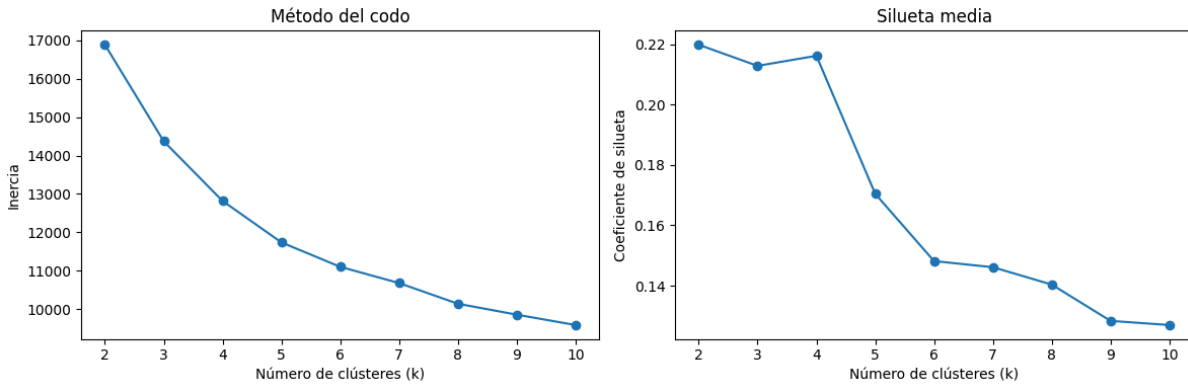


Fig. A.8. Análisis de codo (a) y silueta (b) para el agrupamiento del Dataset 2.

Aunque se exploraron configuraciones con un mayor número de clústeres (por ejemplo, $k = 4$) con el fin de examinar posibles subgrupos más específicos, las métricas de validación y la interpretabilidad de los centros resultantes ratificaron que el modelo bimodal representa de manera más fiel la estructura latente de los datos.

En consecuencia, para cada subconjunto de datos se fijó $k = 2$ como el número óptimo de clústeres así como en [31]. Esta decisión permite dividir a los estudiantes en dos tipologías claramente diferenciadas en función de su sensibilidad emocional y su relación con el rendimiento, tal como se detalla en las secciones siguientes. La elección de dos grupos evita la sobre segmentación que dificultaría la interpretación pedagógica de los perfiles.

A.3.3 Visualización de perfiles mediante componentes principales

En aras de explorar la estructura resultante de los datos y validar visualmente la separación de los dos clústeres obtenidos con el algoritmo *K-means*, se aplicó un análisis de componentes principales (*PCA*, por sus siglas en inglés) sobre las trece características normalizadas que describen a cada estudiante. Esta técnica de reducción de dimensionalidad permite proyectar los individuos en un espacio de menor dimensión (en este caso, dos componentes principales) conservando la máxima varianza posible. La representación gráfica de los clústeres en el plano definido por la Componente Principal 1 (*PC1*) y la Componente Principal 2 (*PC2*) se muestra para los subconjuntos del Dataset 1 en la Fig. A.9 y para los subconjuntos del Dataset 2 en la Fig. A.10.

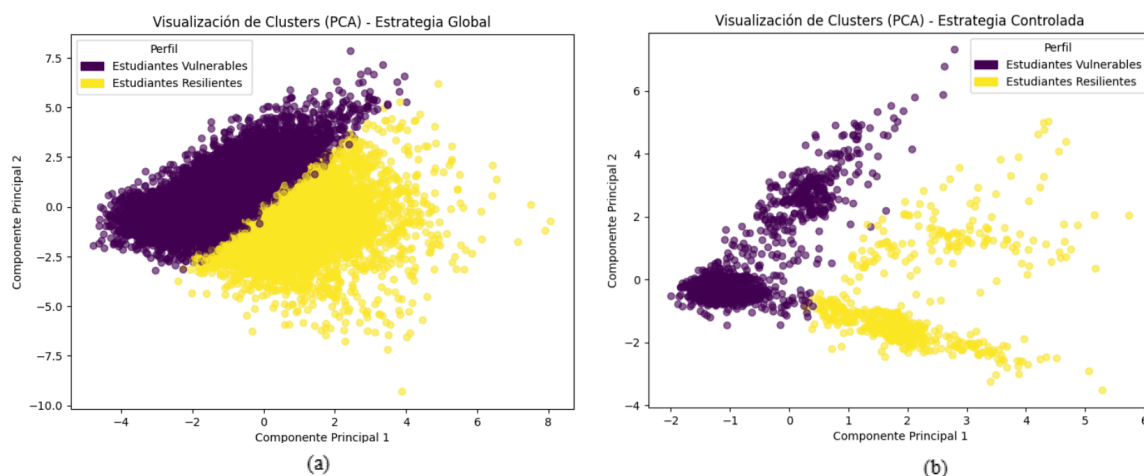


Fig. A.9. Proyección en el plano de PCA de los clústeres de estudiantes del dataset 1 ($k=2$): (a) Agrupamiento para el subconjunto Global y (b) Agrupamiento para el subconjunto Controlado.

En la Fig. A.9 (a) se observa una clara diferenciación entre ambos clústeres a lo largo de la primera componente principal, que explica la mayor parte de la varianza total. Los *estudiantes vulnerables* se ubican predominantemente hacia valores negativos de PC1, formando una nube compacta y de escasa dispersión; mientras que los *estudiantes resilientes* se concentran en la región de valores positivos, mostrando una dispersión ligeramente mayor. La separación es nítida, sin zonas de solapamiento apreciables. En este subconjunto, el clúster 0 al cual denominaremos el grupo de *estudiantes resilientes* presenta una tasa de aciertos media del 71,7 %, correlaciones de frustración y confusión cercanas a cero (0,013 y 0,048 respectivamente) y una variabilidad emocional moderada (ver Tabla A.1). Por su parte, el clúster 1 en el que están los *estudiantes vulnerables*, tiene una tasa de aciertos inferior (64,6 %), correlación negativa para confusión (−0,156) y una variabilidad mucho más baja en confusión ($std = 0,053$). Estos patrones sugieren que el grupo resiliente mantiene un rendimiento alto a pesar de altibajos emocionales, mientras que el vulnerable ve afectado su desempeño cuando aparecen emociones negativas (frustración y confusión).

En la Fig. A.9 (b) se observa el subconjunto Controlado, donde la disposición de los clústeres en el plano PCA es igualmente nítida. El clúster 0 corresponde al perfil de estudiantes resilientes, este tiene mejor rendimiento (77,9 %), correlación positiva de frustración (0,115) y mayor variabilidad emocional ($std = 0,174$). El clúster 1, se asocia con el perfil de *estudiantes vulnerables* ya que presenta una tasa de aciertos del 72,2 % (menor que el clúster 0), una correlación fuertemente negativa entre frustración y acierto (−0,559), y una

variabilidad casi nula en frustración ($std = 0,008$). A pesar de que los valores absolutos de rendimiento son más altos en este subconjunto debido a la selección de las primeras interacciones, la relación entre emociones y éxito sigue los mismos principios: los estudiantes resilientes mantienen o incluso mejoran su rendimiento cuando experimentan emociones negativas, mientras que los vulnerables se desestabilizan fácilmente.

ASSIST2012										
Sub conjunto	clúster	media correct	corr concent	corr frust	corr conf	corr aburr	media frust	media conf	desviación frust	desviación conf
Global	Cl 0	0,717	0,116	0,013	0,048	-0,112	0,389	0,074	0,093	0,222
	Cl 1	0,646	0,240	-0,015	-0,156	-0,165	0,389	0,014	0,081	0,053
Controlado	Cl 0	0,779	0,235	0,115	-0,042	-0,165	0,045	0,045	0,174	0,092
	Cl 1	0,722	0,198	-0,559	-0,111	-0,168	0,043	0,046	0,008	0,089

Tabla A.1. Caracterización de los perfiles estudiantiles según el rendimiento y la correlación emocional (ASSISTments 2012).

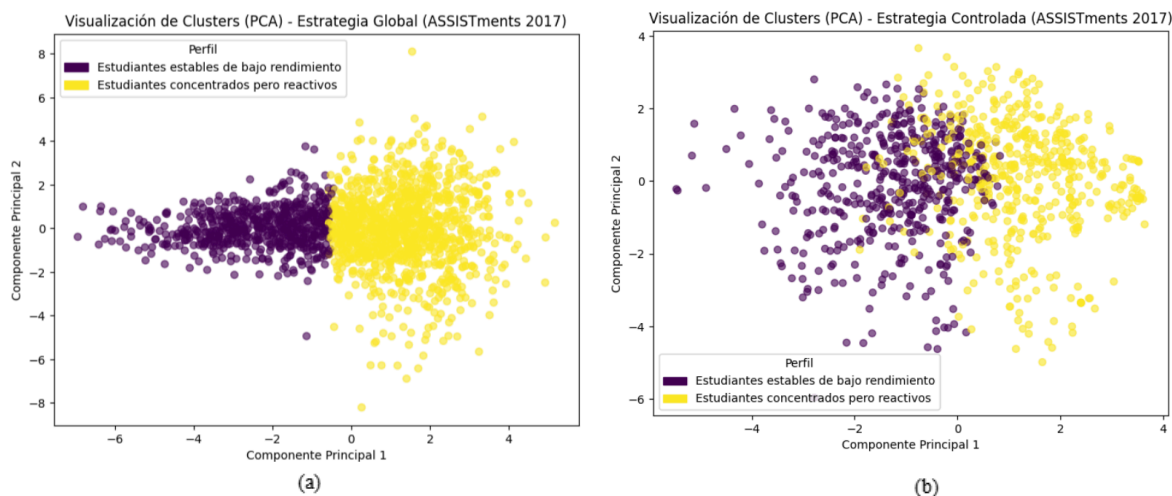


Fig. A.10. Proyección en el plano de PCA de los clústeres de estudiantes del dataset 2 ($k=2$): (a) Agrupamiento para el subconjunto Global y (b) Agrupamiento para el subconjunto Controlado.

Del mismo modo, para el Dataset 2, la Fig. A.10 (a) revela una estructura de separación clara entre ambos clústeres a lo largo de la primera componente principal (PC1). El clúster 0 se define por contener a los estudiantes *estables de bajo rendimiento* estos se

sitúan mayoritariamente en el semiplano negativo de PC1, mientras que el clúster 1 que son los estudiantes *concentrados pero reactivos* ocupan el semiplano positivo. La nube de puntos del clúster 0 presenta una mayor dispersión a lo largo de PC2, lo que indica una heterogeneidad intra-grupo más amplia en comparación con el primer dataset (ASSISTments 2012). En este subconjunto global, el clúster 1 alcanza una tasa de aciertos media del 40,8%, superando claramente al clúster 0, que registra un 30,9%. Esta diferencia invierte la tendencia observada en el dataset ASSISTments 2012, donde el clúster de mayor rendimiento era el resiliente (con correlaciones emocionales cercanas a cero). En cuanto a las emociones, el clúster 1 presenta una concentración media ligeramente inferior (0,634 frente a 0,663) y un aburrimiento medio superior (0,272 frente a 0,219) en comparación con el clúster 0. Las desviaciones estándar son similares en ambos grupos, pero las correlaciones con el rendimiento (Tabla A.2) muestran patrones muy distintos: el clúster 1 tiene una fuerte correlación positiva en cuanto a la concentración con el éxito (0,478) y una fuerte correlación negativa del aburrimiento (-0,563), mientras que en el clúster 0 dichas correlaciones son moderadas (0,299 y -0,316, respectivamente). La confusión también presenta una correlación negativa más marcada que en el clúster 1 (-0,306) que en el clúster 0 (-0,150). Estos resultados indican que los estudiantes *concentrados pero reactivos* son más sensibles a los estados emocionales ya que cuando el estudiante se concentra, su rendimiento se eleva notablemente, pero las emociones negativas lo pueden perjudicar. Por el contrario, los estudiantes *estables de bajo rendimiento* muestran una relación más atenuada entre emociones y aciertos, lo que podría reflejar una menor implicación cognitiva o una mayor dificultad para regular su estado emocional.

Por otra parte, en la Fig. A.10 (b) la proyección PCA mantiene una separación no tan pronunciada pero visible entre ambos clústeres. En el clúster 1 se agrupan los estudiantes *concentrados pero reactivos*, los cuales vuelven a mostrar una tasa de aciertos superior (51,7% frente al 39,7% del clúster 0). Cabe destacar que en este subconjunto el clúster 1 presenta una correlación aparentemente muy alta con la frustración (0,799), pero esta es un artefacto estadístico debido a la varianza prácticamente nula de esa emoción (media de frustración 0,005, desviación típica 0,008), lo que indica que estos estudiantes nunca reportan frustración, por ende, la frustración no es relevante para caracterizar este grupo. En cambio, las correlaciones con confusión (-0,469) y aburrimiento (-0,635) son fuertemente negativas y

sí presentan variabilidad significativa, reforzando el perfil de alta reactividad. El clúster 0 está definido por los estudiantes *estables de bajo rendimiento*, mostrando correlaciones negativas moderadas (confusión $-0,332$, aburrimiento $-0,555$) y una varianza considerable en frustración. Ambos subconjuntos coinciden en que los estudiantes de mayor rendimiento son precisamente los más sensibles a la confusión y al aburrimiento, un patrón opuesto al observado en ASSISTments 2012.

ASSIST2017										
Sub conjunto	clúster	media correct	corr concent	corr frust	corr conf	corr aburr	media frust	media conf	desviación frust	desviación conf
Global	Cl 0	0,309	0,299	0,031	-0,150	-0,316	0,132	0,086	0,297	0,232
	Cl 1	0,408	0,478	0,076	-0,306	-0,563	0,140	0,132	0,299	0,271
Controlado	Cl 0	0,397	0,522	-0,024	-0,332	-0,555	0,208	0,101	0,339	0,178
	Cl 1	0,517	0,502	0,799	-0,469	-0,635	0,005	0,083	0,008	0,132

Tabla A.2. Caracterización de los perfiles estudiantiles según el rendimiento y la correlación emocional (ASSISTments 2017).

En ambos conjuntos de datos (ASSISTments 2012 y 2017), la caracterización estadística y la visualización mediante componentes principales confirman que la partición bimodal ($k = 2$) es la más adecuada, proporcionando perfiles con una interpretación pedagógica coherente. Es fundamental destacar que la concentración es la única dimensión emocional que presenta una correlación positiva dominante y consistente en los clústeres de ambos datasets. Por esta razón, se determinó que la concentración sea la variable principal incorporada para la implementación técnica de los modelos RE-BKT, resaltando también que las otras emociones (frustración, confusión, aburrimiento) se conservaron durante la fase de agrupamiento (clustering), ya que su variabilidad y promedios resultaron determinantes para definir la sensibilidad emocional de cada perfil.

Apéndice B: Evaluación de las curvas de ROC

B.1 Evaluación del clúster 0 (Dataset 1 - Global)

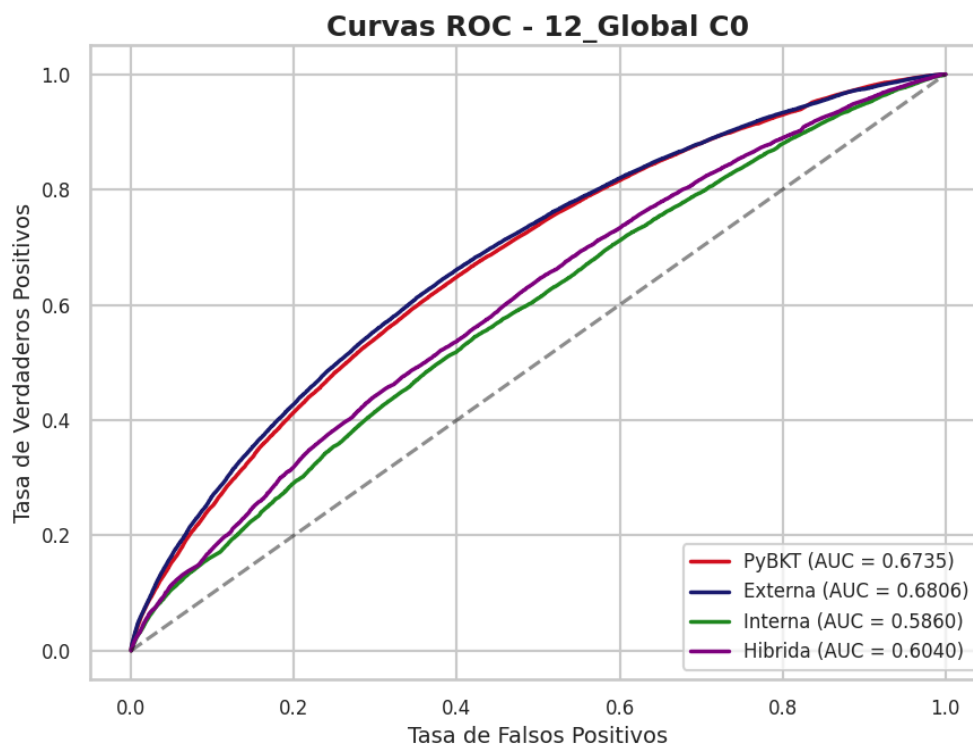


Fig. B.1. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 0).

En la Fig. B.1, correspondiente al subconjunto Global para los *estudiantes resilientes* (clúster 0), se observa que la arquitectura del Enfoque 2 posee el mejor desempeño discriminativo, con un $AUC = 0,6806$ superando ligeramente la referencia de pyBKT ($AUC = 0,6735$). Por el contrario, las arquitecturas del Enfoque 1 ($AUC = 0,5860$) e Híbrida ($AUC = 0,6040$) presentan un rendimiento sustancialmente inferior. Este mayor poder de clasificación sugiere que el modelo Externo logra una identificación más precisa y clara del estado de dominio histórico para estos estudiantes con una base sólida.

B.2 Evaluación del clúster 1 (Dataset 1 - Global)

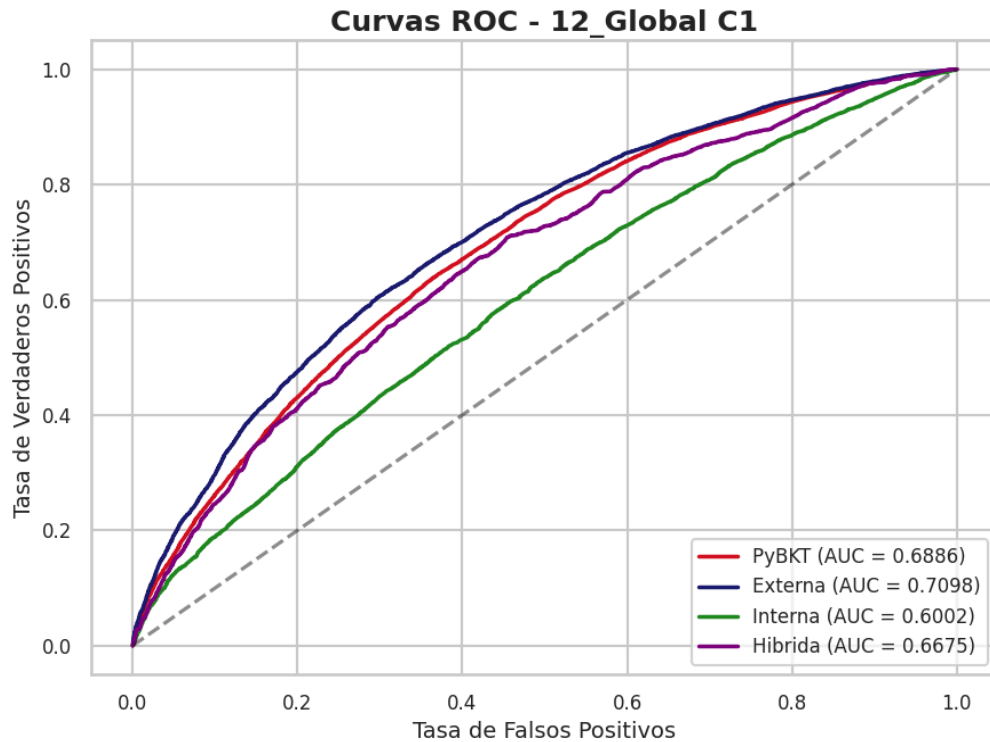


Fig. B.2. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 1).

La Fig. B.2 ilustra el comportamiento predictivo para los *estudiantes vulnerables* (clúster 1) en el escenario histórico. En este grupo, caracterizado por una alta sensibilidad emocional, la arquitectura del Enfoque 2 ratifica su superioridad con un $AUC = 0,7098$, logrando una mejora notable frente al $0,6886$ del modelo base pyBKT. Este incremento en el área bajo la curva confirma que, para estudiantes con alta sensibilidad emocional, la concentración actúa como un diferenciador crítico del desempeño a largo plazo.

B.3 Evaluación del clúster 0 (Dataset 1 - Controlado)

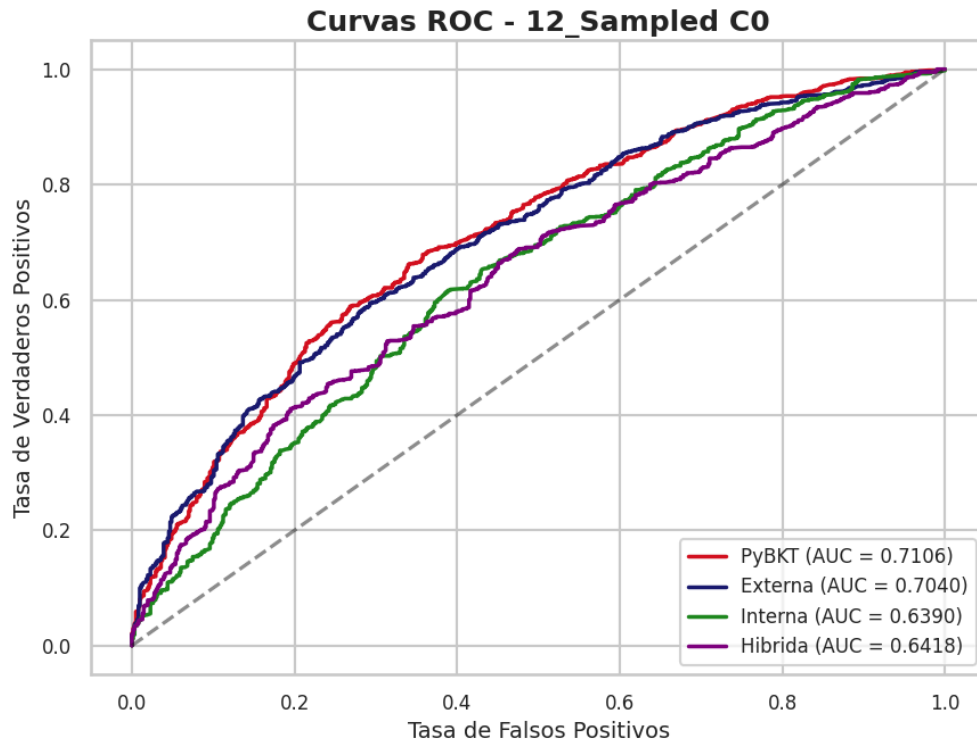


Fig. B.3. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 0).

Los resultados del subconjunto Controlado para el clúster 0 (Fig. B.3) presentan una dinámica atípica donde el modelo base pyBKT conserva una ligera ventaja discriminativa ($AUC = 0,7106$) sobre la arquitectura del Enfoque 2 ($AUC = 0,7040$). Este comportamiento sugiere que, en las primeras 6 interacciones de estudiantes con buen rendimiento, la simplicidad del modelo tradicional es eficiente, mientras que la incorporación de variables emocionales en un lapso de tiempo tan reducido puede introducir ruido estocástico.

B.4 Evaluación del clúster 1 (Dataset 1 - Controlado)

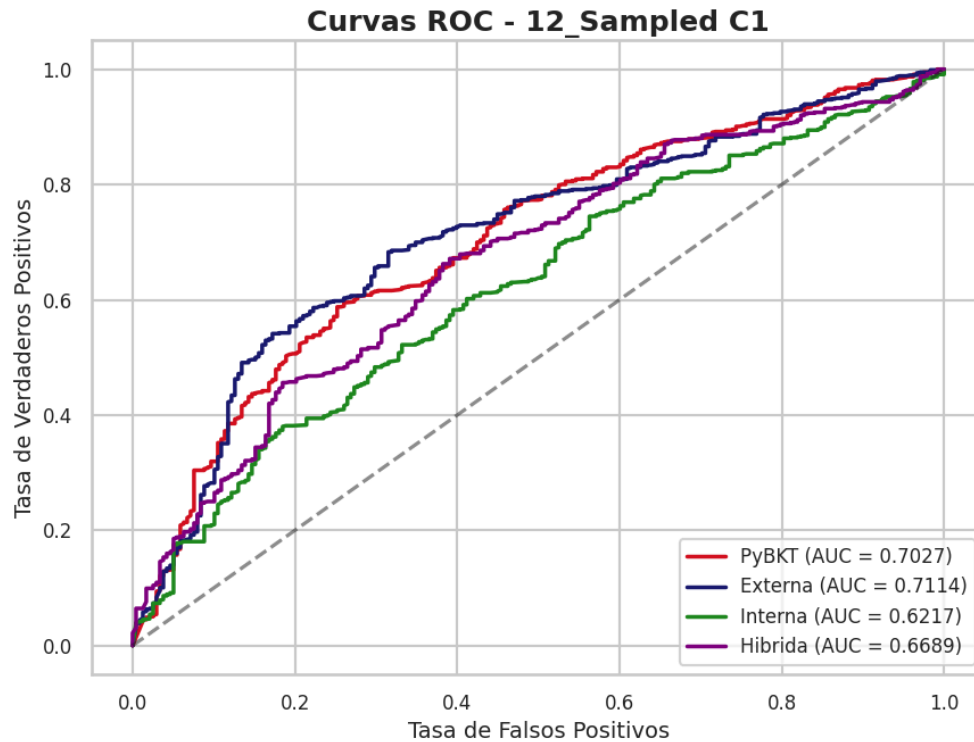


Fig. B.4. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 1).

En este escenario de *estudiantes vulnerables* y datos limitados (Fig. B.4), la arquitectura del Enfoque 2 recupera la superioridad predictiva con un $AUC = 0,7114$, superando el $0,7027$ de pyBKT. Este resultado es de gran relevancia pedagógica, ya que indica que incluso en la fase temprana del aprendizaje, la concentración es un detector confiable para identificar si un estudiante vulnerable está realmente adquiriendo la habilidad o si sus errores responden a desconexión emocional o falta de dominio de la habilidad.

B.5 Evaluación del clúster 0 (Dataset 2 - Global)

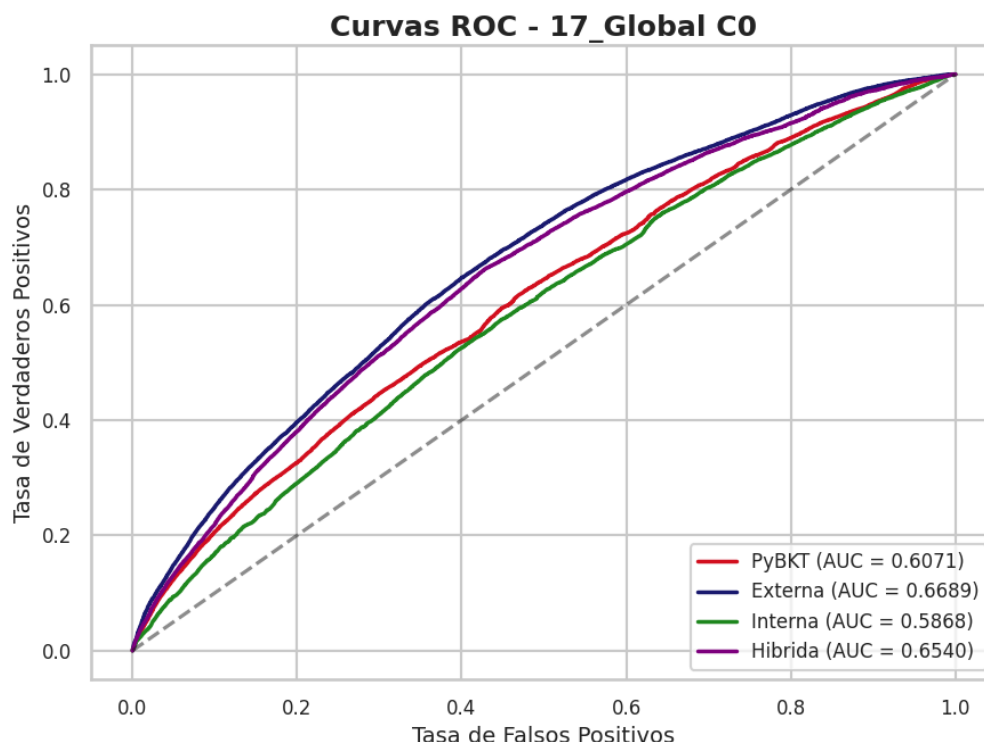


Fig. B.5. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 0).

En la Fig. B.5, correspondiente al subconjunto histórico de estudiantes con desempeño *estables de bajo rendimiento* (clúster 0), la arquitectura del Enfoque 2 domina la capacidad discriminativa con un $AUC = 0,6689$, superando al modelo base pyBKT ($0,6071$). La curva de ROC de la Externa se sitúa consistentemente por encima de la del modelo base indicando una mejora sostenida en la capacidad de discriminar entre aciertos y errores. Las arquitecturas Interna con $AUC = 0,5868$ y la Híbrida con $AUC = 0,6540$ presentan un rendimiento inferior, lo que permite evidenciar que incluso para estudiantes con bajo desempeño, la modulación logística externa de la concentración logra extraer información relevante que el modelo base no considera.

B.6 Evaluación del clúster 1 (Dataset 2 - Global)

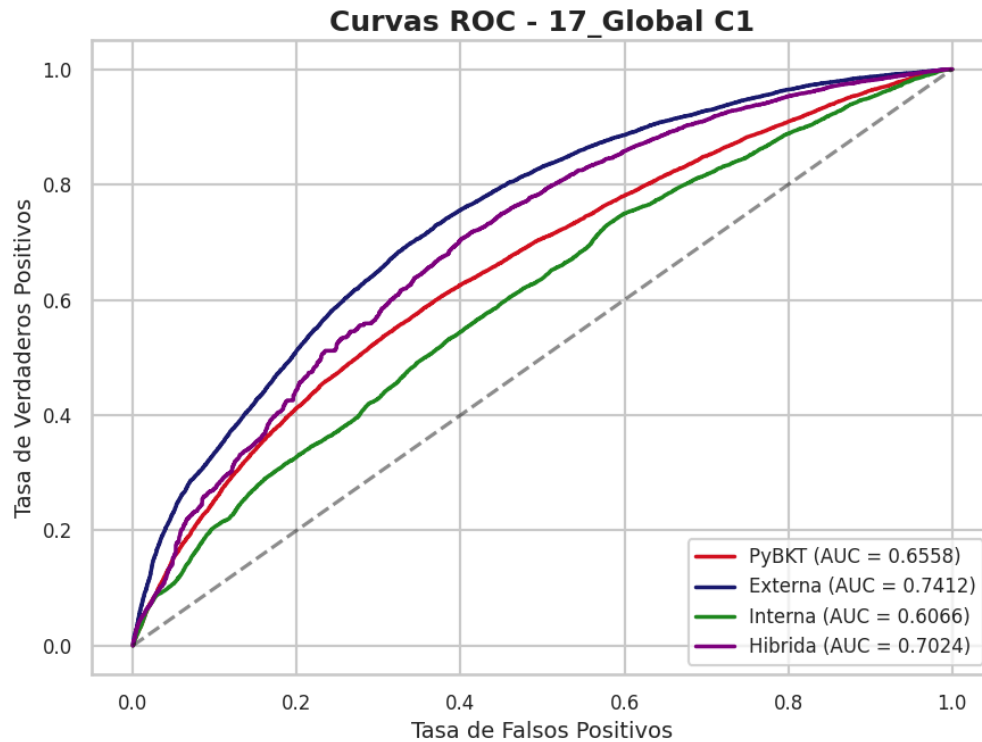


Fig. B.6. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 1).

La Fig. B.6 ilustra los resultados para los estudiantes *concentrados pero reactivos* en el escenario global (clúster 1). Aquí, la arquitectura del Enfoque 2 alcanza su mejor desempeño con un $AUC = 0,7412$, superando el $0,6558$ de pyBKT. La diferencia de $+0,0854$ puntos en el AUC es la más amplia del subconjunto Global de este dataset, lo que demuestra que la incorporación externa de la concentración es particularmente beneficiosa cuando los estudiantes presentan una alta sensibilidad emocional. La arquitectura híbrida también mejora la base ($0,7024$), aunque sin alcanzar el rendimiento de la Externa. Por el contrario, el Enfoque 1 ($0,6066$) degrada el poder discriminativo, confirmando que la integración de la emoción en la matriz de observación es menos efectiva.

B.7 Evaluación del clúster 0 (Dataset 2 - Controlado)

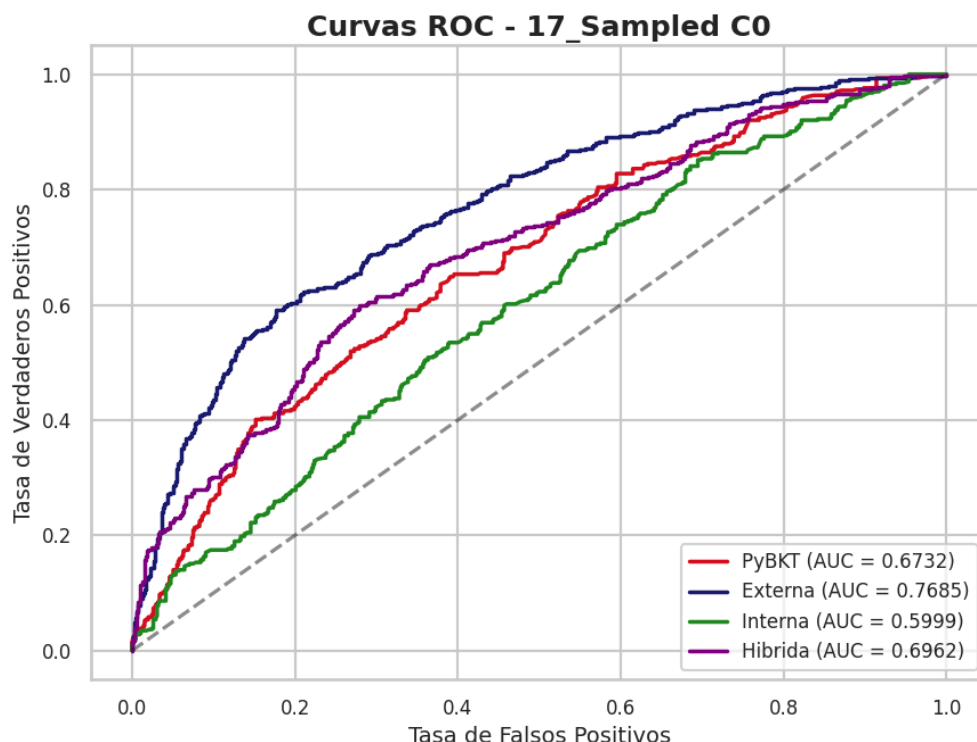


Fig. B.7. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 0).

En el subconjunto Controlado para los estudiantes *estables de bajo rendimiento* (clúster 0), la arquitectura del Enfoque 2 nuevamente lidera con un AUC de 0,7685, superando ampliamente al modelo base (0,6732) y a las demás variantes. La diferencia de +0,0953 puntos en el AUC representa la mejora más sustancial de todo el estudio, lo que indica que incluso con solo seis interacciones por estudiante, la señal de concentración permite al modelo diferenciar con notable precisión a los estudiantes que dominan la habilidad. La arquitectura Híbrida (0,6962) mejora con respecto al base, mientras que la Interna (0,5999) empeora el rendimiento.

B.8 Evaluación del clúster 1 (Dataset 2 - Controlado)

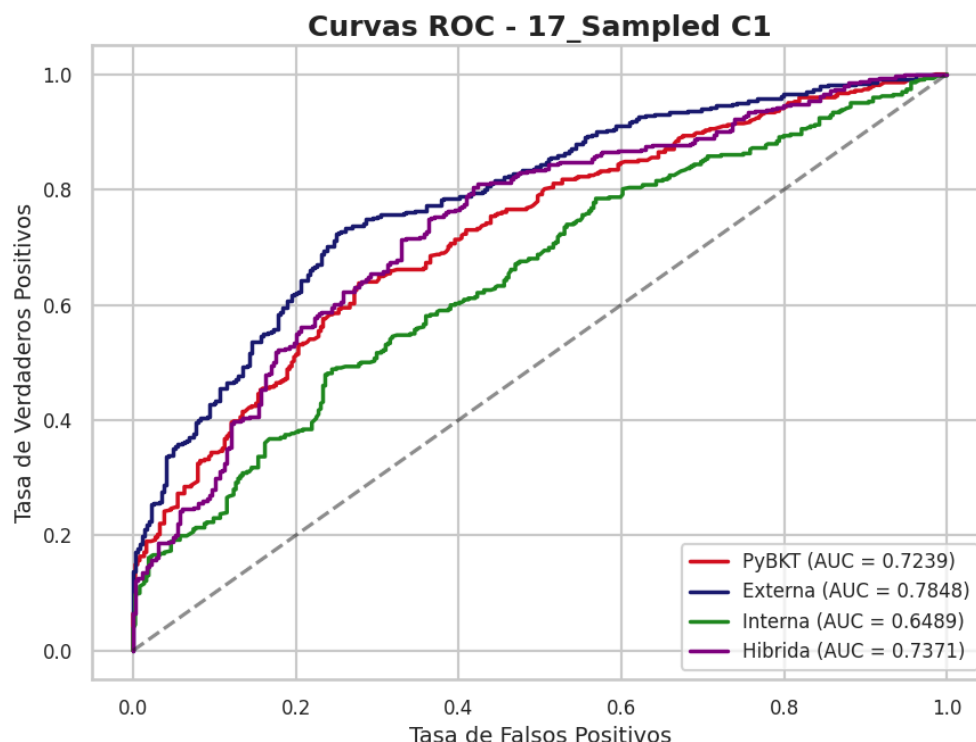


Fig. B.8. Comparación entre el Modelo pyBKT y RE-BKT para la Curva ROC (clúster 1).

Finalmente, en la Fig. B.8, el subconjunto Controlado para los *concentrados pero reactivos* revela que la arquitectura del Enfoque 2 logra un $AUC = 0,7848$, el rendimiento más robusto de toda la investigación, superando al modelo base que tiene $AUC = 0,7239$ y evidentemente a las arquitecturas Interna ($0,6489$) e Híbrida ($0,7371$). La curva ROC correspondiente muestra que la Externa mantiene una tasa de verdaderos positivos superior para cualquier nivel de falsos positivos, lo que confirma su robustez incluso en condiciones de datos limitados. En contraste, el Enfoque 1 presenta el peor desempeño, corroborando que forzar la emoción dentro de la matriz de emisión perjudica la capacidad predictiva, independientemente del perfil estudiantil o del volumen de datos.