

Estudio Comparativo de Múltiples Máquinas

Algoritmos de aprendizaje para el desempeño de los estudiantes

Datos de Inserción Laboral en la Universidad

Athreya Shetty B, Akram Pasha, Amith Singh S, Shreyas NI, Adithya R Hande

Resumen: En la era de la evolución de los datos, muchas organizaciones han tomado la delantera en el almacenamiento de datos en enormes repositorios de datos. El análisis de datos presenta varios desafíos desde el momento en que se capturan los datos hasta que se infieren los conocimientos a partir de los datos. Acentuar la precisión del análisis de datos es de suma importancia ya que muchas decisiones críticas dependen totalmente de los resultados del análisis. El aprendizaje automático se ha encontrado como la herramienta más efectiva y preferida en la literatura para el análisis de datos en memoria. Las universidades recopilan principalmente los datos estadísticos relacionados con los estudiantes que solo se utilizan cuantitativamente o se analizan escasamente para obtener información que podría ser útil para que las autoridades mejoren el porcentaje de ubicaciones en las unidades del campus realizadas a través del análisis temprano de dichos datos con precisión. El trabajo propuesto en este artículo formula el problema de predicción de la probabilidad de que un estudiante sea colocado en una empresa como un problema de clasificación binaria. Luego se esfuerza por entrenar y realizar el estudio empírico de seguir múltiples algoritmos de aprendizaje automático con los datos de ubicación; Regresión logística, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor y Decision Tree. Los modelos de clasificación de aprendizaje automático están diseñados para predecir las probabilidades de que un estudiante sea colocado en una empresa en función de los puntajes académicos, los logros, la experiencia laboral (pasantía) y muchas otras características relevantes del estudiante. Dicho análisis ayuda a las autoridades universitarias a crear dinámicamente planes para mejorar los estudiantes poco probables que se colocarán en una empresa que participa en el reclutamiento del campus que se lleva a cabo en la universidad. Para mejorar estos modelos y evitar que los modelos se sobreajusten a los datos de entrenamiento, se aplican estrategias como la validación cruzada K-Fold para varios valores de k. Los modelos de aprendizaje automático seleccionados también se comparan por su eficiencia mediante el empleo de técnicas de extracción de características supervisadas y no supervisadas, como PCA y LDA. El modelo Decision Tree con K como 10 para la validación cruzada y PCA ha superado a todos los demás modelos produciendo una precisión del 72,83 % con soporte y recuperación satisfactorios durante la experimentación. La aplicación se enfoca en el grupo objetivo de estudiantes, para eventualmente mejorar la probabilidad de que los estudiantes sean colocados durante las campañas de reclutamiento del campus que se llevan a cabo en la universidad.

Términos del índice: minería de datos, análisis de datos, clasificación, Aprendizaje automático, validación cruzada K-Fold, PCA, LDA

I. INTRODUCCIÓN

Existe la necesidad de mejorar las habilidades de los estudiantes que asisten a las prácticas en el campus [1] [2] [3]. El estudio de sus datos previos con algoritmos de aprendizaje automático nos indica las medidas necesarias a tomar para mejorar la calidad de los alumnos y aumentar la probabilidad de ser colocado [4][5][6]. Los métodos que se implementan y utilizan en las universidades son celdas de ubicación donde tienen todos los datos de ubicación de los estudiantes e interactúan con ellos para mejorar las ubicaciones [7][8][9]. Estos datos de ubicación se pueden usar de manera más eficiente con la implementación de modelos de aprendizaje automático para mejorar las ubicaciones, así como para predecir si un estudiante podría ser ubicado o no y las razones relevantes [10][11][12][13][14]. Después de mucha experimentación e investigación, hemos encontrado el modelo de aprendizaje automático más eficiente adecuado para la tarea con

comparación con muchos otros algoritmos de predicción de aprendizaje automático (árbol de decisiones) [15] [16] [17] [18].

Durante el curso de esta experimentación e investigación, hemos seleccionado 5 modelos básicos de aprendizaje automático lineal y no lineal, que son Regresión logística, Naïve Bayes, Máquina de vectores de soporte, K-Vecino más cercano y Árbol de decisiones. Junto a estos modelos, para mejorar el resultado, hemos utilizado K-Fold Cross Validation, PCA y LDA. Hemos calculado la exactitud, la precisión, la recuperación y la puntuación f1 del modelo para respaldar nuestro resultado y la eficiencia del modelo.

Las principales contribuciones del trabajo propuesto son:

- Incorpora todo el ciclo de vida del desarrollo del modelo de aprendizaje automático desde la recolección de datos hasta la visualización de resultados.
- Realiza limpieza de datos ocupándose de valores nulos y valores atípicos.
- Realiza la transformación de los datos para hacerlos adecuados para encajar en un modelo de aprendizaje automático utilizando técnicas de escalado de datos
- Entrena 5 máquinas modelos de aprendizaje.
- Realiza k-fold cross validación sobre 5

Manuscrito revisado recibido el 25 de abril de 2019.

Athreya Shetty B, Escuela de Informática y Tecnología de la Información, Universidad REVA, Bangalore, India.

Akram Pasha, Escuela de Informática y Tecnología de la Información, Universidad REVA, Bangalore, India.

Amith Singh S, Escuela de Informática y Tecnología de la Información, Universidad REVA, Bangalore, India.

Shreyas N I, Escuela de Informática y Tecnología de la Información, Universidad REVA, Bangalore, India.

Adithya R Hande, Escuela de Informática y Tecnología de la Información, Universidad REVA, Bangalore, India.

- Realiza el análisis crítico de desempeño de los 5 modelos de clasificación utilizando varias métricas de evaluación.

La Sección V concluye el documento presentando las direcciones futuras de este trabajo.

En el año 2016, el trabajo de [19], realizó la investigación sobre la predicción de la ubicación de los estudiantes. Los modelos utilizados fueron Regresión Logística, K-NN y SVM. Los atributos clave considerados fueron detalles académicos, habilidades de comunicación, trabajo en equipo y habilidades de programación. En el trabajo de [20], se estudió y presentó la predicción de ubicación basada en el estado psicológico del estudiante. Han utilizado los datos en tiempo real del estudiante de último año con Neural Network y Decision Tree Classifier para la predicción. Se construyó un analizador de colocación de estudiantes en el trabajo de [21], en el que el aprendizaje automático identifica a los estudiantes cuyas habilidades técnicas e interpersonales se buscaban mejorar. En el trabajo de [22] se construyó un modelo de clasificación y predicción para mejorar la precisión de la predicción de ubicación de los estudiantes. Las técnicas de agrupamiento utilizadas fueron canopy, EM, Farthest First, Hierarchical Cluster y SimpleKMeans. Los datos de rendimiento de los estudiantes se utilizaron en el trabajo de [23] para predecir el tipo de empresa en la que encajará el estudiante. Han utilizado el algoritmo Decision Tree y las técnicas de agrupación en clústeres de los datos utilizando herramientas de minería de datos como WEKA. En el año 2017, el trabajo de [24] había presentado un artículo sobre Predicción de Abandono Educativo utilizando técnicas de aprendizaje automático como SVM. Pudieron descubrir los atributos que se requerían para predecir la deserción de los estudiantes.

tercero METODOLOGÍA

modelo de aprendizaje que tiene la mayor precisión de predicción.



análisis para modelos predictivos. Decision Tree divide el conjunto de datos en subconjuntos más pequeños. Puede manejar ambos

datos categóricos y numéricos.

El paso inicial en el desarrollo de los modelos de clasificación es preprocesar [9] los datos, lo que implica la limpieza de datos, el manejo de valores nulos, la conversión de valores de cadena a valores numéricos, la búsqueda de características importantes y muchas más técnicas de preprocesamiento para convertir los datos en un formato comprensible.

Planteamiento del problema:

El modelo de aprendizaje automático utilizado en este trabajo es un problema de clasificación binaria que predice la clase de una nueva observación en función de las observaciones que se entrenaron a un modelo.

Dado que la clase tiene valores binarios {0, 1}, matemáticamente, los modelos de aprendizaje automático se pueden definir de la siguiente manera.

El clasificador de regresión logística crea la hipótesis como se muestra en la ecuación (1) con $h(x)$ = sigmoide (2)

$$Z = WX + B \quad (1)$$

El Clasificador Naïve Bayes crea la hipótesis como se muestra en la ecuación (2)

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right)P(A)}{P(B)} \quad (2)$$

El clasificador de vectores de soporte crea la hipótesis como se muestra en la ecuación (3)

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 = \sum_{i=1}^2 (x_i \cdot y_i) \quad (3)$$

El clasificador de K-vecinos más cercanos crea la hipótesis como se muestra en la ecuación (4)

$$\mathbf{y} = \frac{i}{k} \sum_{i=1}^2 \mathbf{y}_i \quad (4)$$

IV. EXPERIMENTACIÓN Y DISCUSIÓN DE RESULTADOS

Los modelos de clasificación elegidos se desarrollan usando la documentación de scikit-learn v0.20.3 [38] y se desarrollan usando python 3.6.

Los modelos de clasificación tienden a dar resultados sesgados en los conjuntos de datos de entrenamiento. Por lo tanto, para evitar el ajuste excesivo de los datos, incorporamos la validación cruzada k-fold, un método estadístico que divide los conjuntos de datos de entrenamiento y prueba para estimar el rendimiento general de los algoritmos de aprendizaje automático.

Los modelos de aprendizaje automático predictivo categorizan a los estudiantes, ya sea que un estudiante sea colocado o no en la unidad del campus, según el historial de desempeño de los estudiantes hasta el momento. Los datos recopilados contienen varias características relacionadas con un estudiante durante su curso de educación.

Aplicar el algoritmo predictivo de aprendizaje automático sobre los datos recopilados y procesados. Los siguientes datos muestran la

Precisión de diferentes modelos.

Tabla 1: Evaluación de Clasificadores

	LR	Naïve Bayes	SVM	KNN	DT
Precisión	60.5	45.41	48.66	55.42	72.83

La tabla 1 muestra las precisiones de los respectivos modelos LR (Regresión logística), NB (Naïve Bayes), SVM (Máquina de vectores de soporte), KNN (K-Vecino más cercano) y DT (Árbol de decisiones). Muestra que la precisión de Naïve Bayes es la mínima de 45,41 y la precisión de Decision Tree es de **72,83**, la más alta.

La tabla 1 también muestra la precisión de los modelos después de la modificación repetitiva del modelo y los datos. Naïve Bayes es un modelo de aprendizaje automático supervisado probabilístico que se utiliza para cualquier tarea de clasificación que realiza la clasificación en función de las estimaciones de probabilidad mediante el teorema de Bayes. La suposición hecha en este modelo es que las características son independientes. Dado que las características en los conjuntos de datos de ubicación no son independientes, la precisión del modelo Naïve Bayes resulta ser la menor en comparación con el resto. El modelo Naïve Bayes funciona extremadamente bien en un conjunto de datos donde todas las características son independientes. Se utilizan principalmente en el filtrado de spam, análisis de sentimientos, sistema de recomendación, etc. Es rápido y fácil de implementar, pero la mayor desventaja es que requiere que el predictor sea independiente.

El árbol de decisión es un algoritmo de aprendizaje automático supervisado que se usa más en tareas de regresión y clasificación para obtener modelos estables. El rendimiento del árbol de decisiones se encuentra más eficiente en las variables de entrada y salida categóricas y continuas. El algoritmo es completamente invariable y permanece estable al escalado de los datos. El árbol de decisión funciona extremadamente bien cuando los atributos en los conjuntos de datos son de diferentes escalas o una combinación de binario y continuo. El conjunto de datos utilizado en el estudio actual coincide exactamente con el requisito del algoritmo del árbol de decisiones, lo que da como resultado una mayor precisión en comparación con los otros modelos utilizados.

Validación cruzada de K-Fold:

La validación cruzada es un método estadístico que se utiliza principalmente para combatir la compensación de sesgo y varianza que se encuentra en cualquier modelo de aprendizaje automático. Es una estrategia que estima el rendimiento general de un modelo de aprendizaje automático. En este estudio actual, se aplica en varios modelos de aprendizaje automático para comparar el rendimiento de todos los modelos y elegir el mejor modelo que se adapte al conjunto de datos utilizado.

El procedimiento tiene un solo parámetro llamado k que se refiere a la cantidad de conjuntos que se dividirán en un conjunto de datos dado mientras entrenar cualquier modelo de aprendizaje automático. El valor específico de k se elige para dividir los conjuntos de datos en un porcentaje aceptable de divisiones en términos de datos de entrenamiento y prueba.



Tabla 2: Validación K-Cross para K=5

K = 5	LR NOTA	SVM KNN DT			
Precisión y	62.74	44.18	53.52	59.34	63.85
Precisión	80,76	52.5	56.75	63.32	70.27
Recuerdo	86,0	25,99	84.00	78.0	74.0
Puntuación F1	83.29	34.76	67.73	68.89	72.08

La tabla 2 muestra los resultados (Exactitud, Precisión, Recuperación, F1-Score) cuando los modelos de aprendizaje automático se aplican sobre el conjunto de datos con la validación cruzada K-Fold para K=5.

La precisión más alta obtenida es para Decision Tree que es 63.85 y el resultado de precisión más bajo se obtiene para Naïve Bayes que es 44.18.

La tabla 3 muestra los resultados (Exactitud, Precisión, Recuperación, F1-Score) cuando los modelos de aprendizaje automático se aplican sobre el conjunto de datos con la validación cruzada K-Fold para K=10.

La mayor precisión obtenida es para Decision Tree que es 67.63 y el resultado de menor precisión se obtiene para Naïve Bayes que es 41.38.

Tabla 3: Validación K-Cross para K=10

K = 5	LR NOTA	SVM KNN DT			
Precisión y	63.61	41.38	54,72	54.86	67.63
Precisión	69,97	31.66	57.88	63.54	72.55
Recuerdo	83.99	20.0	84.0	70.0	76.0
Puntuación F1	76.34	24.51	68.53	66.61	74.23

Análisis de componentes principales (PCA):

El análisis de componentes principales (PCA) se utiliza como una de las técnicas de extracción de características más utilizadas que reduce la dimensión de los datos sin perder gran parte de la información.

Operaciones matriciales simples de álgebra lineal y estadística se utilizan en este método para calcular una proyección de los datos originales en el mismo número de menos dimensiones.

Tabla 4: ACP, K=5

K = 5	LR NOTA	SVM KNN DT			
Precisión y	58.35	47.82	47.89	55.42	69.78
Precisión	63,38	52.50	56.75	63.32	69.74
Recuerdo	86.00	25,99	84.00	78.00	78.00
Puntuación F1	58.92	40.67	45.42	54.18	64.28

La tabla 4 muestra los resultados (Exactitud, Precisión, Recuperación,

F1-Score) cuando los modelos de aprendizaje automático se aplican sobre el conjunto de datos con la validación cruzada K-Fold para K=5 y Análisis de componentes principales. La mayor precisión obtenida es para Decision Tree que es 69,78 y el resultado de menor precisión se obtiene para Naïve Bayes que es 47,82.

Tabla 5: ACP, K=5

K = 5	LR NOTA	SVM KNN DT			
Precisión y	60.50	45.41	48.66	52.08	72.83
Precisión	63,39	31.66	57.88	63.54	77.78
Recuerdo	84.99	20,00	83.99	70,00	78,00
Puntuación F1	59.91	35.21	46.36	49.63	69.03

La tabla 5 muestra los resultados (Exactitud, Precisión, Recuperación, F1-Score) cuando los modelos de aprendizaje automático se aplican sobre el conjunto de datos con la validación cruzada K-Fold para K=5 y análisis de componentes principales. La precisión más alta obtenida es para el Árbol de decisión que es 72.83 y la menor

El resultado de precisión se obtiene para Naïve Bayes, que es 45,41.

Análisis Discriminante Lineal:

La regresión logística es un algoritmo de clasificación principalmente adecuado para problemas de clasificación binaria. Mientras que el análisis discriminante lineal es una técnica de clasificación lineal preferida cuando el número de valores de decisión en una variable de decisión es más de dos.

Tabla 6: LDA, K = 5

K = 5	LR NOTA	SVM KNN DT			
Precisión y	58,35	47,82	47,53	55.42	63.42
Precisión	63,38	52.50	56.75	63,32	71,86
Recuerdo	86,00	25,94	84.00	78,00	78,00
Puntuación F1	58.92	40.67	45.42	54.18	67.60

La tabla 6 muestra los resultados (Exactitud, Precisión, Recuperación, F1-Score) cuando los modelos de aprendizaje automático se aplican sobre el conjunto de datos con la validación cruzada K-Fold para K=5 y Análisis Discriminante Lineal. La precisión más alta obtenida es para Decision Tree que es 63,42 y el resultado de precisión más bajo se obtiene para SVM que es 47,53.

Tabla 7: LDA, K = 10

K = 5	LR NOTA	SVM KNN DT			
Precisión y	59.50	45.41	48.66	52.08	68.91
Precisión	63.39	31.66	57.88	63.54	77.58
Recuerdo	83.99	20.00	83.99	70.00	76.00
Puntuación F1	59.91	85.21	46.46	49.63	68.07

La tabla 7 muestra los resultados (Exactitud, Precisión, Recuperación, F1-Score) cuando los modelos de aprendizaje automático se aplican sobre el conjunto de datos con la validación cruzada K-Fold para K=10 y análisis discriminante lineal. La mayor precisión obtenida es para Decision Tree que es 68,91 y el resultado de menor precisión se obtiene para Naïve Bayes que es 45,41.

La precisión del modelo depende en gran medida de la gran cantidad de datos recopilados. También para el mejor conjunto de datos con el modelo Random Forest, podemos tener una mejor precisión. Cuando se agregan nuevas características al conjunto de datos, la precisión aumentará automáticamente. Después de modificar el modelo una y otra vez y también al agregar nuevas funciones, puede haber una gran mejora en el rendimiento de los estudiantes, los resultados de ubicación y las predicciones hechas por el modelo. La tendencia general de los estudiantes que fueron colocados fueron los estudiantes con un buen puntaje SGPA, puntaje 10 y 12, incluidas las personas con logros adicionales y experiencia en pasantías.

Se puede proporcionar capacitación adicional a los estudiantes de preúltimo año. Con los datos y recursos disponibles, se puede proporcionar a los alumnos una formación muy particular en lugar de una formación general común a todos. Se pueden realizar clases adicionales para mejorar los puntajes semestrales de los estudiantes para cumplir con los criterios de la compañía. Además, el modelo se utiliza para predecir la razón por la que no se coloca.

El pequeño conjunto de datos y la Entropía en el Árbol de decisión tuvieron un mayor impacto en la obtención de resultados altos para el Árbol de decisión en nuestro trabajo actual. La entropía en el árbol de decisión controla cómo un árbol de decisión decide dividir los datos e impacta inmensamente en cómo un árbol de decisión dibuja sus límites. La entropía en un árbol de decisión se calcula utilizando la ecuación (5), donde $p(X)$ indica la probabilidad de las observaciones en un conjunto de datos.

$$Entropy = - \sum p(X) \log p(X) \dots \dots (5)$$

V. CONCLUSIÓN Y ALCANCE FUTURO

La analítica de datos de ubicación es una de las tareas que se pueden ejercer en el campo de la minería de datos educativos. Este documento recopila los datos de rendimiento de los estudiantes y entrena 5 modelos de clasificación de aprendizaje automático. El resultado de la clasificación del trabajo propuesto nos da una imagen detallada de

el desempeño de los estudiantes que los colocaría. Se encontró que la precisión del algoritmo del árbol de decisiones es de 72,83, que es la más alta de todas y en la mayoría de los casos. El algoritmo con la menor precisión fue Naïve Bayes, que es 45,41 respectivamente. La gran cantidad de estudiantes a nivel universitario, estatal o nacional puede formar los grandes datos. Entonces, el desarrollo de modelos de clasificación de computación distribuida sigue siendo una dirección futura de este trabajo.

RECONOCIMIENTO

Los autores agradecen a la gerencia y al personal de la Escuela de C&IT de la Universidad REVA, Bengaluru, por proporcionar soporte de datos y todas las demás instalaciones para llevar a cabo este proyecto.

REFERENCIAS

- [1] Klopferstein K, Thomas MK. El vínculo entre la experiencia de colocación avanzada y el éxito universitario temprano. *Revista Económica del Sur*. 2009 1 de enero: 873-91.
- [2] Berger DM. Evaluación y ubicación obligatorias: la visión desde un departamento de inglés. *Nuevas direcciones para los colegios comunitarios*. 1997 diciembre; 197 (100): 33-41.
- [3] Morgan DL, Michaelides MP. Establecimiento de puntajes de corte para la colocación universitaria. *Informe de Investigación No. 2005-9. Consejo de Educación Superior*. 2005.
- [4] Clement, A. y Murugavel, T., 2015. Inglés para la empleabilidad: un estudio de caso del análisis de las necesidades de capacitación en inglés para estudiantes de ingeniería en la India. *Enseñanza del idioma inglés*, 8(2), pp.116-125.
- [5] Albraikan A, Hafidh B, El Saddik A. iAware: un sistema de biorretroalimentación emocional en tiempo real basado en señales fisiológicas. *Acceso IEEE*. 2018;6:78780-9.
- [6] Miller, Ryan, Hang Ho, Vivienne Ng, Melissa Tran, Douglas Rappaport, William JA Rappaport, Stewart J. Dandorf, James Dunleavy, Rebecca Viscusi y Richard Amini. "Presentamos un modelo de cadáver fresco para la capacitación en acceso venoso central guiado por ultrasonido en la educación médica de pregrado". *Revista occidental de medicina de emergencia* 17, no. 3 (2016): 362.
- [7] Reeves LM, Schmorow DD, Stanney KM. Cognición aumentada y tecnología de evaluación del estado cognitivo: objetivos de investigación a corto, mediano y largo plazo. *Conferencia internacional sobre los fundamentos de la cognición aumentada* 22 de julio de 2007 (págs. 220-228). Springer, Berlín, Heidelberg.
- [8] Sharma AS, Prince S, Kapoor S, Kumar K. PPS: sistema de predicción de ubicación mediante regresión logística. In 2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE) 2014 Dec 19 (pp. 337-341). IEEE.
- [9] Jeevalatha T, Ananthi N, Kumar DS. Análisis de rendimiento de la selección de ubicación de estudiantes de pregrado utilizando algoritmos de árboles de decisión. *Revista Internacional de Aplicaciones Informáticas*. 2014 1 de enero; 108 (15).
- [10] Kabra RR, Bichkar RS. Predicción del rendimiento de estudiantes de ingeniería utilizando árboles de decisión. *Revista Internacional de Aplicaciones Informáticas*. 2011 diciembre; 36 (11): 8-12.
- [11] Kuzilek Hlosta M, Hørrmannova D, Zdrahal Z, Wolff A. OU Analyse: análisis de estudiantes en riesgo en The Open

Universidad. Revisión de análisis de aprendizaje. 2015 marzo 18:1-6.

PERFIL DEL AUTOR

- [12] Mishra T, Kumar D, Gupta S. Datos de los estudiantes de minería para el rendimiento de la predicción. In2014 Cuarta Conferencia Internacional sobre Tecnologías Avanzadas de Computación y Comunicación 2014 8 de febrero (págs. 255-262). IEEE.



El Sr. Athreya Shetty B está cursando la Licenciatura en Tecnología en la Universidad REVA, Bangalore, India.

- [13] Pandey M, Sharma VK. Un algoritmo de árbol de decisiones relacionado con el análisis y la predicción del desempeño de los estudiantes. Revista Internacional de Aplicaciones Informáticas. 2013 1 de enero; 61 (13).



Prof. Akram Pasha, Profesor Asociado, Escuela de Informática y Tecnología de la Información, Universidad REVA, Bengaluru, India, posee una Licenciatura en Ingeniería de la Universidad de Mysore, una Maestría en Ingeniería de la Universidad de Bangalore y actualmente está cursando un Doctorado en la Universidad Tecnológica de Visveswaraya, Belagaum, India. Sus principales áreas de interés de investigación son las aplicaciones de las técnicas de inteligencia computacional en el análisis de datos.

- [14] Taruna S, Pandey M. Un análisis empírico de las técnicas de clasificación para predecir el rendimiento académico. En 2014 IEEE Internacional Conferencia de Computación Avanzada (IACC) 2014 21 de febrero (págs. 523-528). IEEE.

- [15] Choudhary R, Gianey HK. Revisión exhaustiva sobre algoritmos de aprendizaje automático supervisado. In2017 Conferencia internacional sobre aprendizaje automático y ciencia de datos (MLDS) 2017 14 de diciembre (págs. 37-43). IEEE.



El Sr. Amith Singh S está cursando la Licenciatura en Tecnología en la Universidad REVA, Bangalore, India.

- [16] Kumar RS, KPJK. ANÁLISIS DEL RENDIMIENTO DE LOS ESTUDIANTES BASADO EN CLASIFICACIÓN Y ENFOQUE MAPREDUCE EN BIGDATA. Revista Internacional de Matemática Pura y Aplicada. 2018;118(14):141-8.



El Sr. Shreyas NI está cursando la Licenciatura en Tecnología en la Universidad REVA, Bangalore, India.

- [17] Lakshmanan GT, Li Y, Strom R. Estrategias de colocación para sistemas de flujo de datos a escala de Internet. Computación en Internet IEEE. 2008 noviembre; 12 (6): 50-60.



El Sr. Adithya R Hande está cursando la Licenciatura en Tecnología en la Universidad REVA, Bangalore, India.

- [18] Thangavel SK, Bkaratki PD, Sankar A. Analizador de colocación de estudiantes: un sistema de recomendación que utiliza el aprendizaje automático. En 2017 4ta Conferencia Internacional sobre Sistemas Avanzados de Computación y Comunicación (ICACCS) 2017 6 de enero (págs. 1-5). IEEE.

- [19] Giri A, Bhagavath MV, Pruthvi B, Dubey N. Un sistema de predicción de ubicación que usa el clasificador de k vecinos más cercanos. En 2016 Segunda Conferencia Internacional sobre Computación Cognitiva y Procesamiento de la Información (CCIP) 12 de agosto de 2016 (págs. 1-4). IEEE.

- [20] Halde RR, Deshpande A, Mahajan A. La psicología ayudó a predecir el rendimiento académico mediante el aprendizaje automático. Conferencia internacional IEEE de 2016 sobre tendencias recientes en electrónica, tecnología de la información y la comunicación (RTEICT) 20 de mayo de 2016 (págs. 431-435). IEEE.

- [21] Thangavel SK, Bkaratki PD, Sankar A. Analizador de colocación de estudiantes: un sistema de recomendación que utiliza el aprendizaje automático. En 2017 4ta Conferencia Internacional sobre Sistemas Avanzados de Computación y Comunicación (ICACCS) 2017 6 de enero (págs. 1-5). IEEE.

- [22] Shukla M, Malviya AK. Modelo modificado de clasificación y predicción para mejorar la precisión de la predicción de la ubicación de los estudiantes. Disponible en SSRN 3351006. 12 de marzo de 2019.

- [23] Pruthi K, Bhatia P. Aplicación de la minería de datos para predecir la ubicación de los estudiantes. In2015 International Conference on Green Computing and Internet of Things (ICGCIoT) 8 de octubre de 2015 (págs. 528-533). IEEE.

- [24] Kumar M, Singh AJ, Handa D. Encuesta de literatura sobre la predicción de la deserción escolar. Revista Internacional de Educación e Ingeniería de Gestión. 2017 1 de marzo; 7 (2): 8.

- [25] Documentación de Python 2.7-docs.python.org/2.7 Scikit Learn Machine Aprendizaje en Python- www.scikit-learn.org