

Agente Inteligente para predecir la elección de carrera de estudiantes de bachiller

Dicson Ferney Quimbayo Diaz^{1,2}

1. Pontificia Universidad Javeriana seccional Cali

2.dicson@javerianacali.edu.co

Resumen: La orientación vocacional es un proceso clave para ayudar a los estudiantes a elegir una carrera universitaria adecuada a sus intereses, habilidades y personalidad. En Colombia, existen programas gubernamentales y privados para brindar orientación vocacional, pero estos tienen limitaciones en su accesibilidad y alcance. La tecnología basada en inteligencia artificial puede ser una alternativa viable para apoyar este proceso, ya que puede brindar soluciones personalizadas y accesibles a una población más amplia. El objetivo de la investigación es construir un agente inteligente para predecir la elección de carrera universitaria de estudiantes de bachillerato. Se utilizarán técnicas de aprendizaje automático con el uso de técnicas de machine learning para generar predicciones eficientes. Se utilizará información obtenida a través de las pruebas del saber 11 y saber Pro, que fueron exportadas desde el Instituto Colombiano para la Evaluación de la Educación (ICFES) y contienen información sobre los estudiantes en dos momentos: cuando terminan el bachillerato y cuando han completado al menos el 70% de su programa de educación superior en Colombia. Los resultados obtenidos de diferentes experimentos con las distintas técnicas de inteligencia artificial fue que la técnica de k-vecinos es la que mejor desempeño resultó para la predicción de la carrera universitaria de un estudiante de bachiller. Esta investigación contribuye a la creación de una herramienta para orientadores vocacionales cuenten con ella para facilitar el proceso de guía de la elección de carrera de los estudiantes de bachiller.

Introducción

El proceso orientación vocacional consiste en la recolección de información del estudiante relacionado con sus actitudes, talentos, personalidad e intereses profesionales, esta información se analiza de forma cuantitativa por medio de pruebas vocacionales; además de entrevistas que ayudan a los profesionales en orientación vocacional a darle una guía al estudiante de las posibilidades que tiene en la elección de carrera universitaria.

En un país como Colombia donde solo el 39 % de los graduados de bachiller accede a educación superior, es vital la orientación vocacional. En Colombia se han implementado programas de orientación vocacional como “Buscando Carrera” o “Colombia Aprende” con el propósito de combatir la deserción universitaria en el país, sin embargo, es una estrategia que tiene un impacto reducido debido a que existe una brecha de conectividad a internet según informes del DANE [1], ocasionando dificultad para que los estudiantes pueden acceder a internet para hacer uso de los programas vocacionales brindados por el estado.

Otras estrategias, como la aplicación de pruebas de orientación vocacional, tienen un impacto limitado a nivel poblacional; puesto que, suelen ser un servicio particular, entonces los estudiantes y sus familias deben pagar para acceder al servicio. Por lo que en países de bajos y medianos

ingresos muy pocos bachilleres pueden acceder a asesoría personalizada para la elección de carrera. Por su parte todos, los estudiantes presentan como requisito de grado en Colombia la prueba saber 11, la cual da un puntaje clasificatorio según el desempeño del estudiante.

Dado lo anterior, es necesario contar con alternativas viables y replicables para apoyar a la labor de las instituciones educativas [2]. La Organización de las naciones Unidas (ONU) en el año 2017 realizo la primera “Cumbre de la inteligencia artificial” y propuso a la Inteligencia Artificial (IA) como una herramienta con el potencial de ayudar a las personas a resolver sus problemas, entre estos, la construcción de una educación con calidad [3]; es decir una reforma personalizada a los procesos de enseñanza y acompañamiento de los estudiantes.

La tecnología basada en inteligencia artificial tiene el propósito de proporcionar herramientas educativas para facilitar la obtención de conocimientos, así como la orientación de soluciones alternativas para procesos rutinarios y extensos [4]. Por lo anterior, la implementación de herramientas tecnológicas tiene un impacto relevante en las disciplinas que interactúan con la educación, como lo es la orientación vocacional.

Así se plantea como objetivo construir un agente inteligente para predecir la elección de carrera universitaria de estudiantes de bachillerato, utilizando técnicas de inteligencia artificial. Para ello, se hace un análisis de bases de datos correspondientes a las pruebas del saber 11 y saber Pro. Datos exportados desde el Instituto Colombiano para la Evaluación de la Educación (en adelante ICFES) contienen la información de los estudiantes en 2 momentos: al salir del bachillerato y al cumplir mínimo el 70 % del plan de estudios de un programa de educación superior en Colombia.

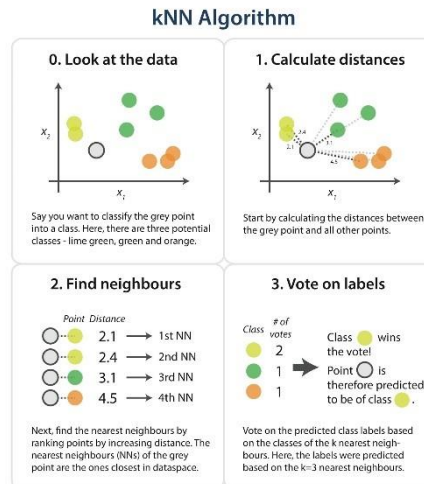
Fundamentación teórica

Para abordar la solución, de un agente inteligente que apoye el proceso vocacional de estudiantes de bachillerato, se utilizan técnicas de aprendizaje automático con técnicas de machine learning. En este sentido, La inteligencia artificial brinda una manera eficiente de generar predicciones, específicamente el aprendizaje automático implica el uso de técnicas de clasificación de patrones con base a un conjunto de datos bajo distintos contextos [4]; adicionalmente el aprendizaje automático tiene las características de ser multidimensional, de procesamiento rápido, de fácil acceso entre otras [3].

El aprendizaje automático, es definido como la capacidad que tiene una computadora de aprender de datos recogidos del ambiente con la intención de mejorarse a sí misma por medio de experiencias previas y estímulos del ambiente donde se encuentre [5]. Ahora bien, dentro de los modelos de predicción de programas de educación superior las técnicas más usadas se amparaban en aprendizaje supervisado, este consiste en que el agente inteligente aprende por medio parejas de datos (entrada y salida), por los cuales se hace una comparación de las salidas del modelo respecto a las salidas esperadas para realizar los arreglos necesarios para mejorar la predicción.

Dentro de las técnicas más relevantes para el proceso se encuentran: K-vecinos. Esta es una técnica de análisis no paramétrica, supervisada de conglomerados y es comúnmente usada en estudios de segmentación. A diferencia de los métodos paramétricos, K-vecinos no asume una distribución subyacente de los datos. Además, con K-vecinos, no existe una fase de entrenamiento explícita. Los cálculos se basan en el conjunto de datos completo en contraste con otras técnicas de aprendizaje automático [6]; en la [figura 1](#) se puede observar el paso a paso del algoritmo para clasificación de las clases.

Figura 1. Pasos del algoritmo K-



vecinos. Tomado de:

<https://www.kdnuggets.com/2016/01/implementing-your-own-knn-using-python.html>

Por su parte, el concepto de orientación vocacional hace alusión a el proceso guiado por un profesional generalmente de psicología, que acompaña a los estudiantes en su elección valorando de forma objetiva un conjunto de aspectos, tales como:

- **Aptitudes:** Habilidades que el estudiante posee y así permite vincularlo a áreas profesionales específicas.
- **Talentos:** habilidades del estudiante, se ha comprobado que quienes trabajan con sus habilidades son personas con un alto nivel de motivación y compromiso.
- **Personalidad:** mide el ajuste socioemocional que tiene el estudiante.
- **Intereses profesionales:** son test relacionados con la orientación vocacional y busca predecir la satisfacción del estudiante en una determinada carrera profesional.

Resultados

Para el presente trabajo se hace un análisis de 857.008 registros pertenecientes a los datos exportados desde el Instituto Colombiano para la Evaluación de la Educación (ICFES). Estos se obtuvieron con el cruce de variables de los resultados del estudiante en 2 momentos: al salir del bachillerato y al cumplir mínimo el 70 % del plan de estudios de un programa de educación superior. rango de años en que se enmarca el conjunto de datos es el siguiente: Saber 11 , 2006 al 2013; Saber Pro, 2012 al 2021.

El presente trabajo contó con la asesoría interdisciplinar de psicólogos y pedagogos que acompañan procesos de orientación vocacional y educativo en el Centro de Enseñanza y aprendizaje de la Javeriana Cali. De acuerdo con la revisión documental y el criterio de los expertos mencionados. Se dio la construcción de la base de datos del modelo predictivo, considerando los aspectos relevantes

para la orientación vocacional como la información socio económica, socio demográfica, resultados de la prueba saber 11, información del colegio, ocupación y nivel académico de los padres y deseos de elección de carrera del estudiante.

Para el entrenamiento de los modelos se debe contar con la claridad de cómo se dividirá el dataset y configuración de cada experimento a realizar. En la [figura 2](#), se muestra da división del conjunto de datos en su parte de entrenamiento y test.

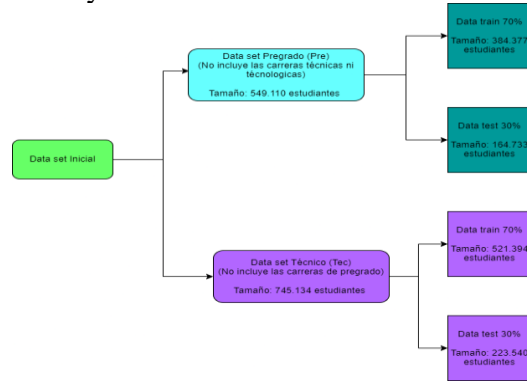


Figura 2. . División del conjunto de datos en los data set's PRE Y TEC

La ejecución de los experimentos se utilizó 2 data sets uno para carrera de profesionales y otro para programas técnicos y tecnológicos. El Test Accuracy fue la métrica para evaluar el desempeño de los modelos creados. En total se realizaron 9 experimentos con modelos y ajustes a nivel de diseño, como de las variables tomadas en cuenta para el proceso de predicción.

De acuerdo con los resultados obtenidos de los experimentos, la técnica de k vecinos presenta el mejor resultado de clasificación de las carreras para ambos datasets como se observa en la [figura 3](#); en el data set profesionales obtuvo un Accuracy de 62 % mientras que en Técnicos y tecnológicos obtuvo 54 %.

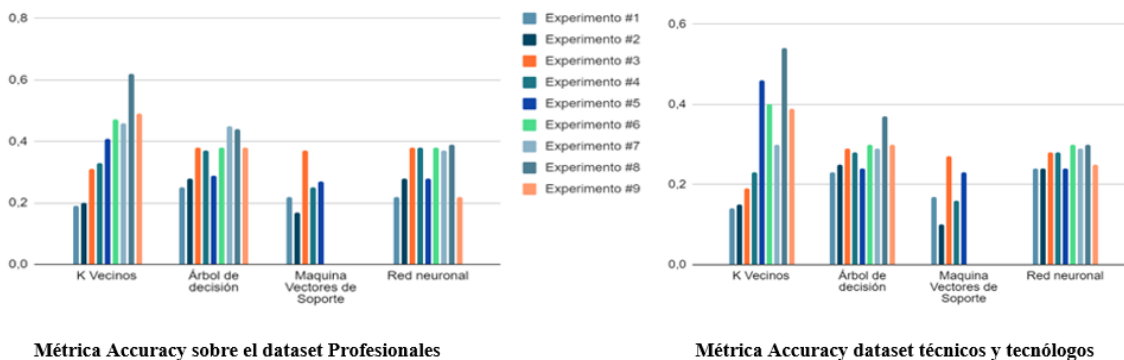


Figura 3. Resultados de la métrica Accuracy de los diferentes experimentos.

Las variaciones de los resultados con el pasar de cada experimento se debe a factores como la cantidad de clases a predecir, balanceo de los datos de entrenamiento, transformación de variables dummies, proporcionando una mejora secuencial en la tarea de la clasificación correcta; ahora una explicación a detalle de la ejecución de estos modelos.

- Disminución de clase: inicialmente la cantidad de clases a predecir eran 11 al realizar el paso de los experimentos, se disminuyó el número de las clases con el objetivo de observar alguna variación en el resultado de la métrica Accuracy; para la reducción en el número de clases se hicieron agrupaciones por área de conocimiento entre los programas de educación superior.
- Balanceo de clases: las variables obtenidas del dataset inicial del ICFES tenían una distribución no paramétrica, algunas clases contaban con datos mayoritarios que superaban miles a otras clases, que contaban con menor número de casos; en consiguiente, no llegaban a establecer una distribución normal para facilitar la técnica la clasificación de la de la variable objetiva predecir. Aun cuando la base datos recopilaba 549.000 (profesional) y 745.134 (Técnico).
- Variables dummies: el data set en su mayoría contaba con variables categóricas éstas, tienen la particularidad que desde las técnicas de machine learning se pueden transformar a variables dummy. Para ello, se hace un arreglo por las n clases que tenga esta variable en la que se representa con un 1 en la casilla donde esté presente el dato, esto se realizó para poder aumentar la diversidad de los datos para las técnicas o predecir.

A partir del modelo de predicción se construyó la interfaz del agente inteligente, El agente inteligente se compone de una interfaz web desarrollada con [streamlit](#) un framework para aplicaciones de machine learning, donde se presenta al usuario los campos a rellenar estilo formulario para ingresar los atributos necesarios para que el modelo proporcione una predicción de la carrera universitaria a estudiar por el recién graduado del bachillerato.

Discusión y conclusiones

La hipótesis inicial fue: “los puntajes del saber 11 relacionados con las áreas del conocimiento eran los suficientemente significativos por sí solos para llegar a predecir la carrera universitaria de un estudiante de bachillerato”. Sin embargo, en la ejecución de los diferentes experimentos se observó que al aumentar variables relacionadas con el contexto del estudiante se obtuvieron resultados mejores en comparación con la hipótesis inicial de la cual se evaluaban solamente los puntajes del saber 11.

Desde la teoría de la orientación vocacional existen 3 ejes para la vocación de un estudiante: los gustos, las aptitudes y el contexto, sin embargo, en este trabajo de grado se enfocó solamente en las aptitudes y datos del contexto recolectados por el ICFES para lograr la predicción de la carrera obteniendo un Accuracy de 0.62 utilizando la técnica k vecinos más cercanos sobre el data set de programas universitarios.

El objetivo es que el orientador vocacional use esta interfaz web para recolectar la información del estudiante para que el agente inteligente pueda realizar la predicción de la carrera universitaria que se adecuada a los datos suministrados, se busca brindar una acompañamiento al orientador vocacional para mejorar los tiempos de atención y volumen de usuarios siendo una herramienta para potenciar el acompañamiento en la toma de decisiones. Teniendo en cuenta aspectos objetivos como lo que brinda el agente inteligente; más la mirada subjetiva de psicólogo en el dialogo con el estudiante bachiller.

Para futuros estudios se recomienda, implementar técnicas de clasificación no supervisada como K-means para evaluar su desempeño y compararlo con el resultado obtenido; realizar el entrenamiento del agente inteligente con datos recolectados propios, con un enfoque en los 3 ejes de la orientación vocacional (gusto, contexto y aptitudes), para complementar los expuestos en repositorios del gobierno y por ultimo explorar la creación de una aplicación móvil o web para que los orientadores vocacionales a nivel nacional hagan uso de esta herramienta.

Finalmente, es importante destacar que la construcción de agentes inteligentes y modelos de predicción en la ingeniería de sistemas puede verse beneficiada y enriquecida con el trabajo interdisciplinario con profesionales de psicología y expertos en orientación vocacional, quienes para este proyecto brindaron una visión diversa al trabajo, lo que permitió obtener una variedad de experimentos ejecutados y contrastados a las hipótesis iniciales.