



TECNOLÓGICO DE ESTUDIOS
SUPERIORES DE COACALCO



UNIDAD DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN
MAESTRIA EN CIENCIAS EN INGENIERIA INDUSTRIAL

TESIS

*MODELO CLASIFICADOR COMO HERRAMIENTA DE
APOYO A LA DECISIÓN EN LA
CONTRATACIÓN DE PERSONAL ACADÉMICO*

QUE PARA OBTENER EL GRADO DE:

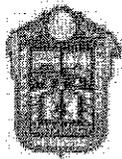
**MAESTRA EN CIENCIAS
EN INGENIERÍA INDUSTRIAL**

LÍNEA DE INVESTIGACIÓN:
ANÁLISIS DE DECISIONES

PRESENTA:
JOSÉ ANTONIO VEGA GONZÁLEZ
102120014

DIRECTOR DE TESIS:
M. EN C. LEOPOLDO VIVEROS ROSAS

COACALCO DE BERRIOZÁBAL AGOSTO 2024



**GOBIERNO DEL
ESTADO DE
MÉXICO**

**Carta Autorización de impresión
de Tesis o Tesina**

FORMATO

FO-TESCo-123

Versión: 4

Pág.: 1 de 3



TESCo
TECNOLÓGICO
DE ESTUDIOS
SUPERIORES
DE COACALCO

**Coacalco, Estado de México a
Agosto 13 del 2024**

CARTA AUTORIZACIÓN DE IMPRESIÓN DE TESIS O TESINA

Nombre del/de la estudiante:

VEGA GONZÁLEZ JOSÉ ANTONIO

**Candidato/a al (posgrado: MAESTRIA EN CIENCIAS DE LA INGENIERÍA INDUSTRIAL
PRESENTE.**

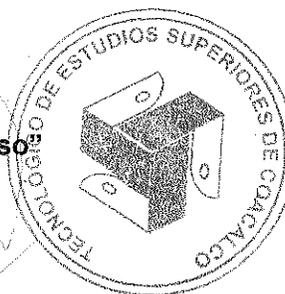
De acuerdo con los Lineamientos para la Operación de los Estudios de Posgrado vigentes en el Tecnológico Nacional de México y habiendo cumplido con todas las indicaciones que el Comité Tutorial le hizo al respecto de su TESIS titulada:

“MODELO CLASIFICADOR COMO HERRAMIENTA DE APOYO A LA DECISIÓN EN LA CONTRATACIÓN DE PERSONAL ACADÉMICO”

Comunicó a usted que esta Subdirección concede su autorización para que proceda a la impresión de la misma.

ATENTAMENTE

“Ciencia, Técnica-Progreso”



MTRO. LEONCIO DAVID ROSADO CRUZ
SUBDIRECCIÓN DE ESTUDIOS PROFESIONALES

TODA COPIA EN PAPEL ES UN “DOCUMENTO NO CONTROLADO” A EXCEPCIÓN DEL ORIGINAL

SECRETARÍA DE EDUCACIÓN
SUBSECRETARÍA DE EDUCACIÓN MEDIA SUPERIOR Y NORMAL
DIRECCIÓN GENERAL DE EDUCACIÓN SUPERIOR
TECNOLÓGICO DE ESTUDIOS SUPERIORES DE COACALCO

AV. 16 DE SEPTIEMBRE No. 54
COACALCO DE BERRIOZÁBAL, MÉXICO
TELS. (0155) 2159-4324, 2159-4325, 2159-4468
www.tecnologicodecoacalco.edu.mx

Agradecimientos

Primeramente, a Dios por mantenerme con vida para llegar a este momento.

Agradezco a todos aquellos que confiaron y me apoyaron para poder terminar esta maestría.

A mi esposa e hijos, que gracias a sus consejos, regaños, ánimos este proceso no sería posible.

A mis amigos, que con su cercanía y apoyo logré avanzar hasta este punto.

Índice

ÍNDICE	1
ÍNDICE DE FIGURAS	3
ÍNDICE DE TABLAS	4
RESUMEN	5
ABSTRACT	6
1. INTRODUCCIÓN	7
2. JUSTIFICACIÓN	12
3. PLANTEAMIENTO DEL PROBLEMA	13
4. ANTECEDENTES DEL PROBLEMA	13
5. OBJETIVOS	14
5.1 Objetivo General	14
5.2 Objetivos Específicos	14
6. HIPÓTESIS	14
7. REVISIÓN DE LA LITERATURA	15
8. MARCO TEÓRICO	28
8.1 Aprendizaje supervisado	28

8.2	Árboles de clasificación	28
8.3	Redes Bayesianas	30
8.3.1	Aprendizaje de redes bayesianas	32
9	DESARROLLO	36
10	RESULTADOS	46
11	CONCLUSIONES Y RECOMENDACIONES	55
12	BIBLIOGRAFÍA	56
13	ANEXOS	59
13.1	Producto Académico 1	59
13.2	Producto Académico 2	64

Índice de Figuras

FIGURA 1. EJEMPLO DE ÁRBOL DE CLASIFICACIÓN.-----	29
FIGURA 2. EJEMPLO DE RED BAYESIANA. -----	30
FIGURA 3. CLASIFICADOR BAYESIANO SIMPLE. -----	32
FIGURA 4. RED DEL CLASIFICADOR NAIVE.-----	34
FIGURA 5. RED DEL CLASIFICADOR NAIVE.-----	37
FIGURA 6. DATOS RECOPIRADOS. -----	39
FIGURA 7. MUESTRA ALEATORIA. -----	40
FIGURA 8. EVALUACIÓN DEL MODELO. -----	52
FIGURA 9. EVALUACIÓN INCORRECTA DEL MODELO. -----	53
FIGURA 10. EVALUACIÓN CORRECTA DEL MODELO. -----	54

Índice de Tablas

TABLA 1. PERSONAL POR CATEGORÍA DEL PERSONAL ACADÉMICO.....	36
TABLA 2. CARACTERÍSTICAS CURRICULARES DEL PERSONAL ACADÉMICO.	37
TABLA 3. CARACTERÍSTICAS CURRICULARES MUESTRA ALEATORIA.	41
TABLA 4. MUESTRA ALEATORIA APLICADO EL CORRECTOR DE LAPLACE.....	42
TABLA 5. PROBABILIDADES CONJUNTAS POR CATEGORÍA.....	43
TABLA 6. PROBABILIDADES INDIVIDUALES POR CATEGORÍA.....	44
TABLA 7. COMBINACIONES DE CARACTERÍSTICAS CURRICULARES.	44
TABLA 7 CONTINUACIÓN. COMBINACIONES DE CARACTERÍSTICAS CURRICULARES.....	45
TABLA 8. PROBABILIDAD POR ALTERNATIVA 1.	47
TABLA 8. PROBABILIDAD POR ALTERNATIVA 64.	49
TABLA 9. PROBABILIDAD 72 ALTERNATIVAS.....	50
TABLA 9 CONTINUACIÓN. PROBABILIDAD 72 ALTERNATIVAS.	51

Resumen

El aprendizaje supervisado se refiere al uso de conjuntos de datos etiquetados para entrenar algoritmos para clasificar datos o predecir resultados. El modelo ajusta sus ponderaciones a medida que se introducen los datos hasta que se haya ajustado adecuadamente como parte del proceso de validación cruzada. Las empresas pueden resolver una variedad de problemas en el mundo real a gran escala con este aprendizaje, en particular cualquier problema de clasificación.

Se utilizan algoritmos de clasificación cuando el resultado es un conjunto infinito de resultados. En este trabajo se utiliza un algoritmo clasificador Naive – Bayes el cual tiene una fuerte suposición de independencia entre los eventos. Los clasificadores de este tipo asumen la presencia o ausencia de una característica particular no se encuentra relacionada con la presencia o ausencia de cualquier otra característica conociendo una variable determinada.

Desarrolla un modelo clasificador Naive - Bayes como propuesta de herramienta para dar claridad en la decisión sobre las alternativas en la contratación de personal académico del Tecnológico de Estudios Superiores de Coacalco (TESCo). Se considera una base de datos de 232 personas como personal académico de las cuales se utilizan 180 para el cálculo de probabilidades, mismas que se prueban con el resto de la base de datos. Se calculan las probabilidades conjuntas y se utilizan para evaluar el clasificador. El modelo permite asignar probabilidades para cada uno de los nuevos prospectos de contratación de personal académico.

Abstract

Supervised learning refers to the use of labeled data sets to train algorithms to classify data or predict outcomes. The model adjusts its weights as data is input until it has been properly tuned as part of the cross-validation process. Companies can solve a variety of large-scale real-world problems with this learning, particularly any classification problem.

Classification algorithms are used when the result is an infinite set of results. In this paper we use a Naive - Bayes classifier algorithm which has a strong assumption of independence between events. Classifiers of this type assume the presence or absence of a particular feature is unrelated to the presence or absence of any other feature knowing a given variable.

A Naive - Bayes classifier model is developed as a proposed tool to provide clarity in the decision about the alternatives in the hiring of academic personnel at the Tecnológico de Estudios Superiores de Coacalco (TESCo). A database of 232 people as academic personnel is considered, of which 180 are used for the calculation of probabilities, which are tested with the rest of the database. The joint probabilities are calculated and used to evaluate the classifier. The model allows assigning probabilities for each of the new academic staff recruitment prospects.

1. Introducción

El Tecnológico de Estudios Supriores de Coacalco (TESCo) fue creado el 4 de septiembre de 1996, teniendo como objetivos [1]

- I. Formar profesionales e investigadores aptos para la aplicación y generación de conocimientos, con capacidad crítica y analítica en la solución de los problemas, con sentido innovador que incorpore los avances científicos y tecnológicos al ejercicio responsable de la profesión de acuerdo con los requerimientos del entorno, el estado y el país;
- II. Realizar investigaciones científicas y tecnológicas que permitan el avance del conocimiento, el desarrollo de la enseñanza tecnológica y el mejor aprovechamiento social de los recursos naturales y materiales que contribuyan a elevar la calidad de vida comunitaria;
- III. Colaborar con los sectores público, privado y social en la consolidación del desarrollo tecnológico y social de la comunidad;
- IV. Realizar programas de vinculación con los sectores público, privado y social que contribuyan a la consolidación del desarrollo tecnológico y social del ser humano;
- V. Realizar el proceso enseñanza-aprendizaje con actividades curriculares debidamente planeadas y ejecutadas; y
- VI. Promover la cultura nacional y universal especialmente la de carácter tecnológico.

De los objetivos de creación del TESCO, se desprende la necesidad de contar con personal académico que coadyuve a lograr los objetivos planteados.

En [2] se define al personal académico como “personal que presta sus servicios para impartir educación superior tecnológica en las asignaturas contenidas en los planes y programas de estudio aprobados por la autoridad educativa federal o, además, para el desarrollo de funciones de investigación, asesoría y apoyo académico. También en [2], se mencionan los requisitos y categorías para el personal académico.

Para ser profesor de asignatura “A” se requiere:

- a) Haber obtenido el título de licenciatura y cédula profesional, expedido por una institución de educación superior pública o con reconocimiento de validez oficial de estudios, otorgado por la autoridad educativa competente, en el área del conocimiento relacionado con la asignatura que se vaya a impartir, y;
- b) Aprobar el examen y haber sido seleccionado en el concurso de oposición respectivo.

Para ser profesor de asignatura “B”, se requiere:

- a) Haber realizado alguna especialización con duración mínima de diez meses en una institución de educación superior o centro de investigación público, o con reconocimiento de validez oficial de estudios otorgado por la autoridad educativa competente; o Contar con al menos dos años de experiencia profesional; o Acreditar al menos dos años como profesor de asignatura “A”, haber cumplido satisfactoriamente sus obligaciones docentes y demostrar su participación en cursos de formación y actualización docente; o Acreditar al menos el grado de maestría, expedido por una institución de educación superior pública, o con reconocimiento de validez oficial de estudios otorgado por la autoridad educativa competente;
- b) Contar con título y cédula profesional;
- c) Aprobar el examen, y haber sido seleccionado, en el concurso de oposición respectivo.

Para ser profesor asociado "A", se requiere:

- a) Tener dos años de haber obtenido título profesional, expedido por una institución de educación superior pública, o con reconocimiento de validez oficial de estudios otorgado por la autoridad educativa competente, así como cédula profesional respectiva; comprobar experiencia docente en nivel superior; o Tener seis años de experiencia profesional en el área de conocimiento correspondiente; y
- b) Aprobar el examen y haber sido seleccionado en el concurso de oposición respectivo.

Para ser profesor asociado "B", se requiere:

- a) Tener cuatro años de experiencia profesional, habiendo desempeñado cargos relacionados con su profesión, y contar con un año de experiencia docente a nivel superior, habiendo aprobado o acreditado su participación en cursos de formación y actualización docente; o Tener el título de licenciatura, y cédula profesional respectiva,

expedido por una institución de educación superior pública, o con reconocimiento de validez oficial de estudios otorgado por la autoridad educativa competente, por lo menos con ocho años de anterioridad a su ingreso o promoción; o Contar con una especialidad realizada en una institución de educación superior pública, o con reconocimiento de validez oficial de estudios otorgado por la autoridad educativa competente, antes de su ingreso o promoción, en un área afín a las asignaturas a impartir; o Tener dos años como profesor de carrera de enseñanza superior asociado "A", haber participado y cumplido con alguna comisión encomendada en el proceso de acreditación de un programa educativo, además, haber participado en cualesquiera de las siguientes actividades: tutorías a estudiantes y pasantes, asesorías en proyectos de extensión y servicio social, residencias profesionales, publicaciones técnico-científicas, tesis, monografías, material didáctico u otros apoyos docentes relacionados con su especialidad como impartición de cursos al sector productivo, cursos al personal del subsistema o de otro subsistema incorporado al modelo de educación superior, haber dictado una conferencia en eventos académicos con reconocimiento local o nacional de nivel superior; y

- b) Aprobar el examen y resultar seleccionado en el concurso de oposición respectivo.

Para ser profesor asociado "C", se requiere:

- a) Acreditar seis años de experiencia profesional, habiendo desempeñado cargos relacionados con su profesión, y contar con dos años de experiencia docente a nivel superior, habiendo aprobado o acreditado su participación en cursos de formación y actualización docente; o Haber obtenido el título y cédula profesional de licenciatura expedido por una institución de educación superior pública, o con reconocimiento de validez oficial de estudios otorgado por la autoridad educativa competente, por lo menos con diez años de anterioridad a su ingreso o promoción académica; o Contar con una especialidad realizada en una Institución de Educación Superior reconocida por la SEP; dos años antes de su ingreso o promoción, en un área afín a las asignaturas a impartir; o Haber obtenido el grado de maestría, y cédula profesional respectiva, expedido por una institución de educación superior pública, o con reconocimiento de validez oficial de estudios otorgado por la autoridad educativa competente; o Tener dos

años como profesor de carrera de enseñanza superior asociado "B", haber participado y cumplido con alguna comisión encomendada en el proceso de acreditación de un programa educativo, además haber participado en cualesquiera de las siguientes actividades: tutorías a estudiantes y pasantes, asesorías en proyectos de extensión y servicio social, residencias profesionales, publicaciones técnico-científicas, tesis, monografías, material didáctico u otros apoyos docentes relacionados con su especialidad como impartición de cursos al sector productivo, cursos al personal del subsistema o de otro subsistema incorporado al modelo de educación superior, haber dictado tres conferencias en eventos académicos con reconocimiento local o nacional de nivel superior; y

b) Aprobar el examen y resultar seleccionado en el concurso de oposición respectivo.

Para ser profesor titular "A", se requiere:

- a) Tener nueve años de experiencia profesional, habiendo desempeñado cargos relacionados con su profesión, y contar con dos años de experiencia docente a nivel superior, habiendo aprobado o acreditado la participación en cursos de formación y actualización docente; o Tener dos años de haber obtenido el grado de maestría, y la cédula profesional respectiva, expedido por una institución de educación superior pública, o con reconocimiento de validez oficial de estudios otorgado por la autoridad educativa competente; o Tener dos años como profesor de carrera de enseñanza superior asociado "C", haber impartido cátedra a nivel superior o de posgrado, contando con publicaciones técnico-científicas y habiendo realizado investigaciones, haber realizado por lo menos tres de las siguientes actividades: elaboración de apuntes, de prototipos, manual de prácticas, impartición de cursos al personal del mismo subsistema o de otro subsistema incorporado al modelo de educación superior, cursos al sector productivo, asesoramiento de tesis, residencia profesional, haber dictado conferencias, haber participado como ponente en simposio, mesas redondas, seminarios, congresos o convenciones con documentos que acrediten la misma;
- b) Haber participado y cumplido con alguna comisión encomendada en el proceso de acreditación de un programa educativo; y
- c) Aprobar el examen y resultar seleccionado en el concurso de oposición respectivo.

Dadas las necesidades actuales el TESCO, se requiere un modelo de clasificación que permite dar claridad en la propuesta de contratación para el nuevo personal académico.

De acuerdo con [3] los tres métodos más usados en aprendizaje automático en tareas de clasificación son redes bayesianas, árboles de decisión y redes neuronales artificiales, además, como lo menciona [4] el objetivo principal de las técnicas de reconocimiento de formas, aplicadas a un problema general de clasificación, consiste en asignar a un objeto o fenómeno físico una de las diversas categorías o clases previamente especificadas.

También [4] indica que las redes bayesianas modelan un fenómeno mediante un conjunto de variables y las relaciones de dependencia entre ellas, con el modelo creado se pueden hacer inferencias bayesianas estimando las probabilidades posteriores de las variables no conocidas, con base a las variables conocidas.

Por su parte [5] mencionan que el clasificador Naive -Bayes, es un modelo efectivo de clasificación por su simplicidad, resistencia al ruido, poco tiempo de procesamiento y alto poder predictivo. Además, se menciona que la tarea del clasificador es etiquetar (clasificar) correctamente un conjunto de datos en uno de los grupos o clases previamente definidas, así, un clasificador bayesiano es una función que asigna a un objeto u observación en la clase con mayor probabilidad

En este trabajo se presenta el desarrollo de un modelo clasificador que permita apoyar la decisión sobre que categoría se le debe asignar al personal académico que se contrate, para ello se hace uso de características deseables por categoría y se usan datos de 232 profesores que forman parte de la plantilla del personal académico del TESCO.

2. Justificación

Reconociendo la Selección de Personal como un proceso de decisión, con base en datos confiables, cuyo fin es el de añadir talentos y competencias que contribuyan al éxito de la organización a largo plazo, además, la tarea de la persona encargada de seleccionar el talento para la empresa ha de ser la de buscar, entre los diferentes candidatos, a aquellos que sean los más adecuados para los puestos de trabajo, o las competencias, que existen en una determinada organización. Dentro de las funciones del personal académico se encuentran: impartir educación para formar profesionales de nivel y medio superiores e investigadores; organizar y realizar investigaciones sobre problemas de interés regional y nacional; desarrollar actividades orientadas a extender los beneficios de la ciencia, la técnica y la cultura, así como participar en la dirección y administración de las actividades mencionadas que la autoridad respectiva le encomiende. Siempre buscando un equilibrio entre Docencia – Investigación y Gestión.

Derivado de la rotación de personal presente dentro de cualquier organización, siempre es recomendable contar con una herramienta que permita dar claridad en lo referente a las decisiones sobre la categoría que se le debe asignar a una nueva contratación de personal académico, siempre conservando principios de equidad y no discriminación.

3. Planteamiento del Problema

La contratación de docentes en los diferentes programas educativos coadyuba al alcance de nuestros objetivos, cumpliendo lo marcado en nuestra misión y visión institucionales.

4. Antecedentes del Problema

La contratación de aspirantes a ocupar un puesto docente dentro del Tecnológico de Estudios Superiores de Coacalco se realiza de acuerdo con lineamientos del TecNM y subprocesos del Sistema de Gestión de Calidad, llevándose a cabo de la siguiente manera:

Al momento de comenzar a elaborarse los horarios correspondientes al siguiente semestre, cada jefatura de división verifica que con los docentes con los que cuenta puedan cubrir las materias a ofertar en dicho semestre, en caso de que con ellos no lo logre, se realiza una publicación de solicitud de puesto en la cual se especifica el grado requerido, los años de experiencia profesional, la experiencia docente y demás requisitos con los cuales se podrán postular todos los aspirantes, después de haberse postulado y dejado sus datos en la pagina o medio donde se publicó la vacante, se procede a contactar al aspirante para realizar una entrevista en la cual se confirmará lo solicitado en la publicación revisando el CV en ese mismo momento se mostrará el horario requerido, si el aspirante y la jefatura no se tiene problema alguno en la revisión de información se procede a programar una clase muestra, la cual será considerada en base a alguna de las asignaturas ofertadas para el puesto y para dicha clase se notificará a docentes de la academia indicándoles que se realizará una clase muestra al aspirante y para lo cual se requiere su presencia, el día propuesto para la clase muestra al aspirante con el fin de que pueda demostrar conocimiento y desenvolvimiento frente a una audiencia se le lleva al grupo donde se requiere su participación, en dicha clase se evaluarán dos puntos en especial que son sus habilidades docentes y las competencias disciplinarias dentro de estos puntos se evaluarán los conocimientos que tiene respecto a la asignatura y como se desarrolla en el manejo de actividades frente a grupo.

Ya realizada esta actividad se llevará a cabo la evaluación del aspirante de acuerdo a lo registrado por los docentes y estudiantes en el reporte de clase muestra con el fin de que se cumple con lo requerido para el puesto; en caso de cumplir y tener buen resultado de dicha evaluación se contactará y notificará al aspirante el resultado y se le solicitará al aspirante reunir los documentos

que son requeridos para su contratación; llegado este momento se realiza el horario que el aspirante cubrirá y deberá firmarlo el aspirante, el jefe de división y dirección académica, dicho horario se entregará junto con el oficio de contratación dirigido al Departamento de Personal.

Una vez que el aspirante ha entregado la documentación al departamento de personal y se ha cumplido con dicha actividad y firmado el contrato, ya se considera docente adscrito a la jefatura que lo solicitó.

En este momento se da por cerrado el proceso de contratación.

5. Objetivos

5.1 Objetivo General

Desarrollar un modelo que permita dar claridad de acción en la decisión de la categoría asignada a personal académico de nueva contratación en el Tecnológico de Estudios Supriores de Coacalco.

5.2 Objetivos Específicos

- Establecer categorías de personal académico con características curriculares.
- Calcular probabilidades condicionales de acuerdo con las características curriculares.
- Desarrollar modelo Naive – Bayes como apoyo a la decisión.

6. Hipótesis

Es posible desarrollar un modelo Naive – Bayes que permita dar claridad en la decisión sobre la categoría que se le debe asignar a una nueva contratación de personal académico del Tecnológico de Estudios Superiores de Cocalco.

7. Revisión de la literatura

La referencia [6] se centra en la aplicación del algoritmo de árbol de decisión en el ámbito del marketing móvil. El marketing móvil ha adquirido una importancia significativa en la era digital, y los algoritmos de aprendizaje automático, como el árbol de decisión, han demostrado ser útiles para analizar y tomar decisiones basadas en datos en este campo. El objetivo de este estudio es analizar cómo el algoritmo de árbol de decisión puede ayudar a las empresas de marketing móvil a comprender mejor el comportamiento del consumidor y a tomar decisiones estratégicas más efectivas. Se comienza proporcionando una visión general del marketing móvil y los desafíos asociados con la toma de decisiones en este ámbito. Luego, se presenta el algoritmo de árbol de decisión y se explican sus conceptos y metodología. A continuación, se describen diferentes aplicaciones del algoritmo de árbol de decisión en marketing móvil, como la segmentación de clientes, la personalización de mensajes y la predicción de la respuesta del consumidor. Se presentan casos de estudio y resultados de investigaciones anteriores para respaldar la efectividad del algoritmo en estas áreas.

También aborda los desafíos y limitaciones del uso del algoritmo de árbol de decisión en marketing móvil, como el manejo de grandes volúmenes de datos y la interpretación de los resultados. Se discuten posibles soluciones y enfoques para superar estas limitaciones, como la combinación de diferentes algoritmos de aprendizaje automático y la utilización de técnicas de visualización de datos.

La referencia [7] centra su estudio en la clasificación de texto en árabe utilizando la técnica de maestro-esclavos. Se menciona que la clasificación de texto es una tarea fundamental en el procesamiento del lenguaje natural, y con el crecimiento exponencial de los datos en árabe, se requieren enfoques eficientes para clasificar y organizar estos textos. La técnica de maestro-esclavos propuesta en este estudio ofrece una solución escalable y paralela para abordar esta problemática. Se describe cómo se implementa la técnica de maestro-esclavos para la clasificación de texto en árabe. El proceso comienza con la selección de un maestro que se encarga de coordinar y distribuir la tarea de clasificación a varios esclavos. Cada esclavo realiza la clasificación en una porción del conjunto de datos y luego envía los resultados al maestro. El maestro combina y

fusiona los resultados de los esclavos para obtener la clasificación final. También se proporciona una visión general de las técnicas de preprocesamiento y extracción de características utilizadas en la clasificación de texto en árabe. Se discuten enfoques como la eliminación de “stop Words”, la normalización de texto y la representación vectorial de los documentos. Estas técnicas se integran en el proceso de clasificación de maestro-esclavos para mejorar la precisión y eficiencia de la clasificación.

Se presentan resultados experimentales que demuestran la efectividad de la técnica de maestro-esclavos en la clasificación de texto en árabe. Se comparan los resultados con otros enfoques de clasificación populares y se muestra que la técnica propuesta logra una precisión destacada. Además, se discuten los beneficios de escalabilidad y paralelismo de la técnica de maestro-esclavos, lo que la hace adecuada para grandes volúmenes de datos en árabe.

En la referencia [8] se muestra la aplicación del método del árbol de decisión para predecir el rendimiento académico de los estudiantes a partir de los valores obtenidos en el último semestre. La predicción del rendimiento estudiantil es de gran importancia para identificar y apoyar a los estudiantes en riesgo de bajo rendimiento. El método del árbol de decisión proporciona una herramienta eficaz para el análisis y la toma de decisiones basadas en datos en este contexto. Se destaca cómo se implementa el método del árbol de decisión para predecir el rendimiento académico de los estudiantes. Se utilizan los valores obtenidos en el último semestre, como calificaciones, promedio ponderado y otros indicadores relevantes, como variables de entrada para construir el árbol de decisión. El árbol de decisión desarrollado se entrena y se ajusta utilizando un conjunto de datos de muestra que incluye información histórica de estudiantes anteriores. Luego, se utiliza el árbol de decisión entrenado para predecir el rendimiento académico de nuevos estudiantes en función de sus valores finales del semestre.

Se presentan casos de aplicación y resultados experimentales que demuestran la eficacia del método del árbol de decisión en la predicción del rendimiento estudiantil. Se comparan los resultados de las predicciones con los valores reales y se evalúa la precisión del método. Además, se discuten las ventajas y limitaciones del enfoque del árbol de decisión en la predicción del rendimiento estudiantil, como la interpretación de los resultados y la necesidad de datos históricos fiables.

Por otra parte, la referencia [9] muestra el uso de árboles de decisión como herramienta en el diagnóstico médico. Refieren que los árboles de decisión son una técnica de aprendizaje automático que ha demostrado ser efectiva en la toma de decisiones basadas en datos en diversos campos, incluida la medicina. En el contexto médico, los árboles de decisión pueden ayudar a los profesionales de la salud a realizar diagnósticos precisos y tomar decisiones clínicas informadas. Se describe cómo se aplican los árboles de decisión en el diagnóstico médico. Estos árboles se construyen utilizando variables relevantes, como síntomas, signos clínicos y resultados de pruebas, como nodos de decisión. A través de una serie de preguntas y condiciones, el árbol guía al médico hacia una clasificación o diagnóstico específico. La estructura de árbol permite una evaluación sistemática de los datos del paciente y la identificación de las condiciones o enfermedades más probables. Se destacan las ventajas de utilizar árboles de decisión en el diagnóstico médico. Estas ventajas incluyen su capacidad para manejar grandes conjuntos de datos, su interpretabilidad, la facilidad para incluir nuevas variables y su eficiencia en términos de tiempo y recursos computacionales. Además, se discuten las limitaciones y desafíos asociados con el uso de árboles de decisión en medicina, como la necesidad de datos de alta calidad y la sensibilidad a cambios en los criterios de clasificación.

Se presentan ejemplos y resultados de investigaciones previas que respaldan la eficacia de los árboles de decisión en el diagnóstico médico. Estos ejemplos abarcan diferentes áreas médicas, como la detección de enfermedades cardíacas, el diagnóstico de cáncer y la predicción de riesgo de complicaciones en pacientes. Los resultados demuestran que los árboles de decisión pueden mejorar la precisión diagnóstica y ayudar a los médicos a tomar decisiones fundamentadas.

La referencia [10] se centra en el uso de árboles de decisión mejorados para estimar “redshifts” fotométricos, una medida crucial en la astrofísica que indica la distancia de un objeto astronómico según su espectro de luz. El método propuesto, llamado ArborZ, utiliza árboles de decisión potenciados para realizar estimaciones precisas de “redshifts fotométricos” a partir de datos obtenidos a través de observaciones fotométricas. Se destaca cómo se implementa ArborZ en el proceso de estimación de “redshifts fotométricos”. Los árboles de decisión mejorados se construyen utilizando una técnica de impulso que combina múltiples árboles de decisión débiles para mejorar la precisión y la capacidad de generalización. Estos árboles se entrenan utilizando datos espectrofotométricos de referencia que contienen información sobre “redshifts” conocidos.

Una vez entrenado, el modelo ArborZ se utiliza para predecir los “redshifts” de nuevos objetos astronómicos a partir de sus observaciones fotométricas.

En el estudio se presentan resultados experimentales que demuestran la eficacia de ArborZ en la estimación de “redshifts fotométricos”. Se comparan las predicciones de ArborZ con mediciones espectroscópicas reales y se evalúa la precisión del método. Los resultados muestran que ArborZ logra una alta precisión en la estimación de “redshifts fotométricos”, lo que indica su utilidad y aplicabilidad en la astrofísica. También, el estudio destaca las ventajas de utilizar árboles de decisión mejorados en la estimación de “redshifts fotométricos”. Estas ventajas incluyen la capacidad de manejar grandes volúmenes de datos, la interpretabilidad del modelo resultante y la eficiencia en términos de tiempo de ejecución. También se discuten las limitaciones y desafíos asociados con el uso de ArborZ, como la calidad de los datos fotométricos y la disponibilidad de datos espectroscópicos de referencia.

En la referencia [3] se menciona que los clasificadores bayesianos son un conjunto de algoritmos de aprendizaje automático que se basan en el teorema de Bayes para realizar clasificaciones. Estos clasificadores son ampliamente utilizados en diversas áreas, como el procesamiento del lenguaje natural, la minería de datos y la clasificación de documentos. Se dice también que, entre los clasificadores bayesianos, el algoritmo Naive Bayes es uno de los más populares debido a su simplicidad y eficiencia. Se describe el enfoque general de los clasificadores bayesianos y se destaca el algoritmo Naive Bayes como uno de los ejemplos más conocidos dentro de esta categoría.

El algoritmo Naive Bayes se basa en el supuesto de independencia condicional entre las características o atributos de los datos. Este supuesto simplifica el cálculo de las probabilidades condicionales y permite una clasificación rápida y eficiente. También se explora cómo se aplica el algoritmo Naive Bayes en la clasificación de datos. El proceso comienza con la construcción de un modelo de entrenamiento utilizando un conjunto de datos etiquetados. Se calculan las probabilidades de cada clase y las probabilidades condicionales de los atributos dadas las clases. Luego, se utiliza el modelo entrenado para clasificar nuevos datos, calculando la probabilidad de pertenencia a cada clase y seleccionando la clase con la probabilidad más alta como la etiqueta de clasificación. Se presentan ejemplos y resultados experimentales que demuestran la efectividad del algoritmo Naive Bayes en la clasificación de datos. Estos ejemplos abarcan diversas

aplicaciones, como la clasificación de correo electrónico como spam o no spam, la detección de enfermedades en diagnósticos médicos y la clasificación de opiniones en el análisis de sentimientos. Los resultados muestran altos niveles de precisión y eficiencia en la clasificación realizada por el algoritmo Naive Bayes.

La referencia [11] se centra en el desarrollo de un clasificador de fuentes de rayos X utilizando el algoritmo Naive Bayes. La clasificación de fuentes de rayos X es una tarea importante en la astronomía de rayos X, ya que permite identificar y categorizar diferentes tipos de objetos astronómicos en función de sus propiedades observadas en la región de rayos X del espectro electromagnético. Se describe cómo se utiliza el algoritmo Naive Bayes para clasificar fuentes de rayos X. El clasificador se entrena utilizando un conjunto de datos etiquetados que contienen información sobre características observadas, como la intensidad, el espectro y la variabilidad de las fuentes de rayos X. Estas características se utilizan como variables de entrada para construir el modelo de clasificación basado en el teorema de Bayes y el supuesto de independencia condicional entre las características.

Se presentan casos de aplicación y resultados experimentales que demuestran la eficacia del clasificador Naive Bayes en la clasificación de fuentes de rayos X. Se comparan las clasificaciones realizadas por el clasificador con clasificaciones realizadas por expertos astrónomos, y se evalúa la precisión y la confiabilidad del clasificador. Los resultados muestran altos niveles de concordancia entre las clasificaciones realizadas por el clasificador Naive Bayes y las clasificaciones de expertos, lo que demuestra su utilidad en la identificación de diferentes tipos de fuentes de rayos X. También se discuten las ventajas y limitaciones del enfoque de clasificación basado en Naive Bayes en el contexto de fuentes de rayos X. Entre las ventajas se incluyen la simplicidad y eficiencia computacional del algoritmo, así como la capacidad de manejar grandes conjuntos de datos. Sin embargo, se resalta la importancia de considerar las limitaciones del supuesto de independencia condicional y la necesidad de tener en cuenta otros enfoques y técnicas complementarias para mejorar la precisión de la clasificación.

En la referencia [12] se enfoca en la clasificación de intercorrelaciones de índices de contaminantes atmosféricos (API, por sus siglas en inglés) utilizando algoritmos de árboles de decisión. La clasificación precisa de las intercorrelaciones de los API es crucial para comprender y abordar los problemas de calidad del aire y sus efectos en la salud humana y el medio ambiente.

Se describe cómo se aplica el algoritmo de árbol de decisión para clasificar las intercorrelaciones de los API. Se utilizan datos recopilados de múltiples estaciones de monitoreo de la calidad del aire, que incluyen mediciones de diferentes contaminantes atmosféricos como dióxido de azufre (SO₂), monóxido de carbono (CO) y partículas suspendidas (PM₁₀). Estos datos se utilizan para entrenar el modelo de árbol de decisión, que clasifica las intercorrelaciones de los API en categorías predefinidas, como "alta correlación", "correlación moderada" y "baja correlación". Se presentan resultados experimentales que demuestran la eficacia del algoritmo de árbol de decisión en la clasificación de las intercorrelaciones de los API. Se evalúa la precisión del modelo utilizando técnicas de validación cruzada y se comparan las clasificaciones obtenidas por el modelo con clasificaciones realizadas por expertos en calidad del aire. Los resultados muestran una alta concordancia entre las clasificaciones realizadas por el algoritmo de árbol de decisión y las clasificaciones de expertos, lo que indica su capacidad para identificar patrones de intercorrelación significativos entre los contaminantes atmosféricos.

Se discuten las ventajas de utilizar algoritmos de árboles de decisión en la clasificación de las intercorrelaciones de los API. Estas ventajas incluyen la capacidad de manejar múltiples variables de entrada, la interpretabilidad del modelo resultante y la capacidad de lidiar con datos faltantes o ruidosos. También se mencionan las limitaciones y desafíos asociados con el uso de este enfoque, como la necesidad de contar con datos de calidad y representativos, así como la interpretación adecuada de los resultados

En la referencia [13] se centran en el uso del análisis de árbol de decisiones para determinar los factores que influyen en el cumplimiento del tiempo de entrega de los pedidos de cotización. El cumplimiento oportuno de los plazos de entrega es crucial para el éxito de las operaciones comerciales, y comprender los factores que afectan esta métrica puede ayudar a las empresas a optimizar sus procesos y mejorar la satisfacción del cliente. Se describe cómo se aplica el enfoque de análisis de árbol de decisiones para determinar los factores que influyen en el cumplimiento del tiempo de entrega de los pedidos de cotización. El análisis de árbol de decisiones se basa en la construcción de un árbol que representa las diferentes variables y condiciones que afectan el tiempo de entrega. Se utiliza un conjunto de datos que incluye información relevante, como el tipo de producto, la disponibilidad de materiales, la capacidad de producción y otros factores relacionados con la entrega.

Se presentan resultados experimentales que demuestran la eficacia del análisis de árbol de decisiones en la identificación de los factores críticos que afectan el cumplimiento del tiempo de entrega. Mediante la construcción del árbol de decisiones, se identifican las variables más influyentes y se determinan las relaciones y dependencias entre ellas. Esto permite a las empresas comprender mejor los factores que pueden causar retrasos en la entrega de los pedidos de cotización y tomar medidas correctivas para mejorar la eficiencia y la puntualidad. Se discuten las ventajas de utilizar el enfoque de análisis de árbol de decisiones en la determinación de los factores que afectan el cumplimiento del tiempo de entrega. Estas ventajas incluyen la capacidad de visualizar de manera clara y concisa las relaciones causales, la facilidad de interpretación de los resultados y la flexibilidad para incorporar nuevos factores o condiciones en el análisis. También se mencionan las limitaciones y consideraciones importantes, como la necesidad de contar con datos confiables y representativos, y la importancia de realizar un análisis continuo y adaptativo a medida que cambian las condiciones del negocio.

En la referencia [14] se muestra el uso del algoritmo de árbol de decisión C4.5 para la clasificación de los beneficiarios de becas familiares. La clasificación precisa de los destinatarios de las becas es esencial para garantizar una asignación justa y eficiente de los recursos destinados a la educación. El algoritmo C4.5 es ampliamente utilizado en la minería de datos y ofrece una solución eficaz para la clasificación basada en árboles de decisión.

Se describe cómo se aplica el algoritmo de árbol de decisión C4.5 para clasificar a los beneficiarios de becas familiares. El algoritmo se basa en la construcción de un árbol de decisión que utiliza características relevantes relacionadas con los solicitantes, como el nivel de ingresos familiar, el rendimiento académico, el tamaño de la familia y otros factores relevantes. El árbol de decisión generado permite clasificar a los solicitantes en categorías predefinidas, como "elegible para beca", "no elegible" o "requiere una revisión adicional".

Se presentan resultados experimentales que demuestran la eficacia del algoritmo de árbol de decisión C4.5 en la clasificación de los beneficiarios de becas familiares. Se utiliza un conjunto de datos etiquetados que incluye información sobre los solicitantes y su elegibilidad para recibir una beca familiar. El algoritmo C4.5 se entrena utilizando estos datos y se evalúa su rendimiento mediante técnicas de validación cruzada y medidas de precisión y exhaustividad. Los resultados muestran altos niveles de precisión y rendimiento en la clasificación de los beneficiarios de becas,

lo que indica la utilidad del algoritmo C4.5 en este contexto. También se discuten las ventajas y limitaciones del uso del algoritmo de árbol de decisión C4.5 en la clasificación de los beneficiarios de becas familiares. Entre las ventajas se incluyen la facilidad de interpretación del árbol de decisión resultante, la capacidad de manejar datos faltantes o ruidosos, y la eficiencia computacional del algoritmo. Sin embargo, se resalta la importancia de contar con datos de calidad y representativos, así como la necesidad de tener en cuenta otros factores relevantes en el proceso de clasificación.

En la referencia [15] se utiliza el clasificador Naive Bayes para predecir el peso del bebé en casos donde la madre padece hipertensión durante el embarazo. La predicción precisa del peso del bebé es esencial para evaluar la salud y el desarrollo fetal, especialmente en casos de hipertensión materna que pueden influir en el crecimiento del feto. El clasificador Naive Bayes ofrece una solución eficaz para realizar esta predicción basada en características relevantes del embarazo y la madre. Se describe cómo se aplica el clasificador Naive Bayes para predecir el peso del bebé de madres con hipertensión. Se utilizan datos clínicos y demográficos recopilados durante el embarazo, como la edad de la madre, el índice de masa corporal, la presión arterial, los niveles de glucosa y otros factores relevantes. Estas características se utilizan como variables de entrada para entrenar el clasificador Naive Bayes, que estima la probabilidad de que el bebé tenga un peso específico dentro de una categoría predefinida (por ejemplo, peso bajo, peso normal, peso alto). Se presentan resultados experimentales que demuestran la eficacia del clasificador Naive Bayes en la predicción del peso del bebé en casos de hipertensión materna. Se evalúa el rendimiento del modelo utilizando técnicas de validación cruzada y medidas de precisión y error. Los resultados muestran una precisión aceptable en la predicción del peso del bebé, lo que respalda la utilidad del clasificador Naive Bayes en este contexto.

También se discuten las ventajas de utilizar el clasificador Naive Bayes para la predicción del peso del bebé de madres con hipertensión. Estas ventajas incluyen su simplicidad, su capacidad para manejar datos faltantes o ruidosos, y su eficiencia computacional. También se mencionan las limitaciones y consideraciones importantes, como la necesidad de contar con datos representativos y actualizados, y la importancia de considerar otros factores clínicos relevantes en la predicción del peso del bebé.

En la referencia [16] se centran en el uso de árboles de decisión para reconocer las actividades diarias humanas. El reconocimiento preciso de las actividades que realizamos a diario es importante en áreas como la salud, el monitoreo del estado físico y el desarrollo de aplicaciones de asistencia personalizada. Los árboles de decisión proporcionan un enfoque efectivo para el reconocimiento de actividades basado en características extraídas de los datos recopilados. Se describe cómo se aplica el algoritmo de árbol de decisión para reconocer actividades diarias humanas. Se recopilan datos sobre movimientos y patrones corporales de individuos mientras realizan diversas actividades, como caminar, correr, sentarse, etc. Estos datos se utilizan para entrenar el árbol de decisión, que clasifica nuevas instancias de datos en las diferentes actividades predefinidas. Se presentan resultados experimentales que demuestran la eficacia de los árboles de decisión en el reconocimiento de actividades diarias humanas. Se utilizan técnicas de evaluación, como validación cruzada, para medir la precisión y el rendimiento del modelo. Los resultados muestran altos niveles de precisión en la clasificación de actividades, lo que respalda la utilidad de los árboles de decisión en este contexto.

Además, se discuten las ventajas y desafíos asociados con el uso de árboles de decisión en el reconocimiento de actividades diarias. Entre las ventajas se incluyen la facilidad de interpretación del árbol resultante, la capacidad para manejar características y datos de diferentes modalidades, y la capacidad de adaptación a diferentes contextos y usuarios. Sin embargo, también se mencionan desafíos como la necesidad de contar con conjuntos de datos representativos y la selección adecuada de características relevantes.

En la referencia[17] se muestra el análisis de sentimiento en redes sociales en relación con las elecciones. El análisis de sentimiento permite comprender las opiniones y actitudes de los usuarios con relación a un tema específico, en este caso, las elecciones. El método Naive Bayes se utiliza como una técnica efectiva para clasificar los mensajes de las redes sociales en categorías de sentimiento, como positivo, negativo o neutro. Se describe cómo se aplica el método Naive Bayes para analizar el sentimiento en redes sociales sobre las elecciones. Se recopilan mensajes de plataformas de redes sociales, como Twitter (ahora X) o Facebook, que contienen información relacionada con las elecciones. Estos mensajes se procesan y se les asigna una etiqueta de sentimiento utilizando el algoritmo Naive Bayes, que se basa en la probabilidad condicional de las palabras con relación a cada categoría de sentimiento.

El estudio presenta resultados experimentales que demuestran la eficacia del método Naive Bayes en el análisis de sentimiento en redes sociales sobre elecciones. Se evalúa el rendimiento del modelo utilizando técnicas de evaluación, como la precisión y la exhaustividad. Los resultados muestran una capacidad satisfactoria del método para clasificar los mensajes en las categorías de sentimiento, lo que respalda su utilidad en este contexto. Además, se discuten las ventajas del método Naive Bayes en el análisis de sentimiento en redes sociales. Estas ventajas incluyen su simplicidad, su capacidad para manejar grandes volúmenes de datos y su eficiencia computacional. También se mencionan las limitaciones y consideraciones importantes, como la necesidad de contar con datos de entrenamiento representativos y la posible presencia de sesgos o ambigüedades en el análisis de sentimiento.

En la referencia [18] se enfocan en la predicción de los clubes de fútbol que podrían ser relegados en la Serie A italiana utilizando el algoritmo de árbol de decisión basado en C4.5. La capacidad de predecir qué clubes tienen más probabilidades de ser relegados es de gran interés tanto para los aficionados como para los profesionales del fútbol, ya que puede ayudar en la toma de decisiones estratégicas y en la planificación de las temporadas futuras. Se describe cómo se aplica el algoritmo de árbol de decisión basado en C4.5 para predecir los clubes relegados en la Serie A italiana. Se recopilan datos históricos de varias temporadas anteriores, como estadísticas de rendimiento, resultados de partidos, clasificaciones y otros factores relevantes. Estos datos se utilizan para entrenar el árbol de decisión, que utiliza las características de los clubes para predecir la probabilidad de ser relegados.

Además, se presentan resultados experimentales que demuestran la eficacia del algoritmo de árbol de decisión basado en C4.5 en la predicción de los clubes relegados en la Serie A italiana. Se evalúa el rendimiento del modelo utilizando medidas de precisión, recall y F1-score. Los resultados muestran una capacidad aceptable para identificar los clubes que tienen más probabilidades de ser relegados, lo que respalda la utilidad del algoritmo en este contexto. De igual forma se discuten las ventajas de utilizar el algoritmo de árbol de decisión basado en C4.5 en la predicción de clubes relegados. Estas ventajas incluyen su capacidad para manejar conjuntos de datos complejos y variables no lineales, su interpretabilidad y su capacidad para identificar las características más relevantes para la predicción. También se mencionan las limitaciones y consideraciones importantes, como la necesidad de contar con datos actualizados y la importancia

de considerar otros factores no incluidos en el estudio, como cambios en el cuerpo técnico o fichajes de jugadores.

En la referencia [19] se muestra el uso de la técnica la técnica Feature Dependent Naive Bayes (Naive Bayes dependiente de características) para la detección de intrusiones en redes. Los sistemas de detección de intrusos son herramientas cruciales para garantizar la seguridad de las redes y proteger la información confidencial. En este estudio, se propone una mejora del algoritmo Naive Bayes mediante la consideración de las características dependientes de las variables de entrada. Se describe cómo se aplica la técnica Feature Dependent Naive Bayes en el contexto de los sistemas de detección de intrusos en redes. Se recopilan datos de tráfico de red, que incluyen información sobre paquetes, direcciones IP, puertos y otros atributos relevantes. Estos datos se utilizan para entrenar el modelo de Naive Bayes mejorado, que tiene en cuenta las dependencias entre las características para mejorar la precisión en la detección de intrusos. Se presentan resultados experimentales que demuestran la eficacia de la técnica Feature Dependent Naive Bayes en la detección de intrusos en redes. Se evalúa el rendimiento del modelo utilizando medidas de precisión, recall y F1-score. Los resultados muestran una mejora significativa en la precisión y la capacidad de detección en comparación con el enfoque tradicional de Naive Bayes, lo que respalda la utilidad de esta técnica en el ámbito de la seguridad de redes.

También se discuten las ventajas y desafíos asociados con el uso de Feature Dependent Naive Bayes en los sistemas de detección de intrusos en redes. Entre las ventajas se incluyen la simplicidad del algoritmo, la eficiencia computacional y la capacidad de adaptación a diferentes tipos de datos de tráfico de red. Sin embargo, también se mencionan desafíos como la necesidad de contar con datos de entrenamiento representativos y la selección adecuada de las características dependientes para cada tipo de intrusión.

En la referencia [20] se muestra la evaluación del riesgo de los clientes antes de otorgarles un préstamo utilizando el algoritmo de árbol de decisión. La evaluación del riesgo es una parte fundamental del proceso de otorgamiento de préstamos, ya que ayuda a las instituciones financieras a determinar la probabilidad de que los clientes incumplan con sus obligaciones de pago. En este estudio, se utiliza el algoritmo de árbol de decisión para construir un modelo de evaluación de riesgo basado en características y datos históricos de los clientes. Se describe cómo se aplica el algoritmo de árbol de decisión en el contexto de la evaluación del riesgo de clientes

antes de otorgar un préstamo. Se recopilan datos relevantes, como la información financiera de los clientes, historial crediticio, nivel de endeudamiento y otros factores pertinentes. Estos datos se utilizan para entrenar el árbol de decisión, que clasifica a los clientes en categorías de riesgo, como bajo riesgo, riesgo moderado o alto riesgo, en función de las características y datos proporcionados. Se presentan resultados experimentales que demuestran la eficacia del algoritmo de árbol de decisión en la evaluación del riesgo de clientes antes de otorgar un préstamo. Se evalúa el rendimiento del modelo utilizando medidas de precisión, recall y F1-score. Los resultados muestran una capacidad satisfactoria del modelo para clasificar correctamente a los clientes en diferentes categorías de riesgo, lo que respalda la utilidad de este enfoque en el ámbito financiero. Además, se discuten las ventajas y limitaciones asociadas con el uso del algoritmo de árbol de decisión en la evaluación del riesgo de clientes. Entre las ventajas se incluyen la interpretabilidad del modelo, la capacidad para manejar datos no lineales y la eficiencia computacional. Sin embargo, también se mencionan desafíos, como la necesidad de contar con datos de entrenamiento representativos y la consideración de otros factores contextuales y económicos que pueden afectar la evaluación del riesgo.

En la referencia [21] se utiliza el algoritmo Naive Bayes para el análisis de sentimiento en Twitter (ahora X). El análisis de sentimiento en redes sociales, como Twitter, es una tarea importante para comprender las opiniones y emociones de los usuarios. En este estudio, se utiliza el algoritmo Naive Bayes, un clasificador probabilístico, para identificar el sentimiento positivo, negativo o neutral de los mensajes de Twitter. Se describe cómo se aplica el algoritmo Naive Bayes en el análisis de sentimiento en Twitter. Se recopilan datos de tweets, que contienen texto y etiquetas de sentimiento, como positivo, negativo o neutral. Estos datos se utilizan para entrenar el modelo Naive Bayes, que utiliza la probabilidad condicional para asignar el sentimiento más probable a nuevos mensajes de Twitter. El estudio presenta resultados experimentales que demuestran la eficacia del algoritmo Naive Bayes en el análisis de sentimiento en Twitter. Se evalúa el rendimiento del modelo utilizando medidas de precisión, recall y F1-score. Los resultados muestran una capacidad satisfactoria del modelo para clasificar correctamente los sentimientos en los tweets, lo que respalda la utilidad de este enfoque en el análisis de sentimiento en redes sociales.

Se discuten las ventajas y limitaciones asociadas con el uso del algoritmo Naive Bayes en el análisis de sentimiento en Twitter. Entre las ventajas se incluyen la simplicidad y eficiencia computacional del algoritmo, su capacidad para manejar grandes volúmenes de datos y su capacidad para adaptarse a diferentes contextos y lenguajes. Sin embargo, también se mencionan desafíos, como la necesidad de contar con datos de entrenamiento bien etiquetados y la sensibilidad a la presencia de palabras ambiguas o sarcásticas en los tweets.

8. Marco Teórico

8.1 Aprendizaje supervisado

Se reconoce que el aprendizaje supervisado permite a las organizaciones resolver una amplia variedad de problemas del mundo real a escala como, utilizando un conjunto de datos de entrenamiento para enseñar a los modelos la salida deseada. El conjunto de entrenamiento debe incluir datos de entrada y resultados correctos permitiendo que el modelo aprenda con el tiempo. Estos algoritmos miden su precisión a través de la función de pérdida de la información buscando siempre la minimización del error [22].

Los algoritmos más comunes para el aprendizaje supervisado son [22]:

1. Redes neuronales
2. Naive bayes
3. Regresión lineal
4. Regresión logística
5. Máquinas de vectores de soporte (SVM)
6. K vecino más cercano (KNN)
7. Bosque aleatorio

8.2 Árboles de clasificación

Los árboles de clasificación entran en dentro de los métodos de clasificación supervisada donde se tiene una variable dependiente o clase, y el objetivo del clasificador es averiguar o determinar dicha clase para casos nuevos. La construcción de estos árboles es mediante un proceso de inducción también denominado *Top-Down-Induction-Decisión-Trees* (TDIDT) [23].

Un clasificador puede ser definido como una función $d(x)$ definida en el espacio de clasificación X , el cual relaciona a cada patrón x del espacio de clasificación con una clase del conjunto de posibles valores a los que puede pertenecer; también se dice que un clasificador es una partición del espacio de clasificación X en M subconjuntos distintos $A_1, A_2, A_3, \dots, A_M$, siendo X la unión de todos ellos y para todo x que pertenece a A_M la clase predicha es C_M [23].

En la Figura 1 se muestra un árbol de clasificación simple.

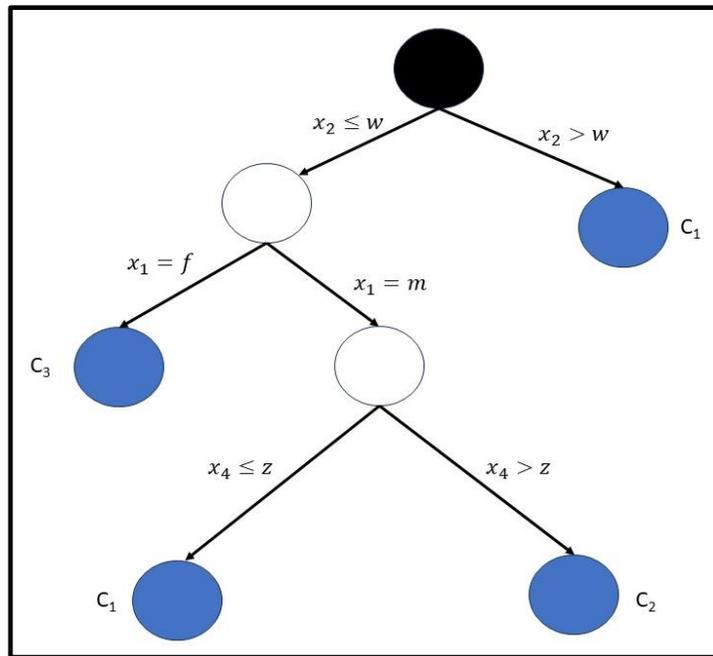


Figura 1. Ejemplo de árbol de clasificación.
Referencia [23]

Como se puede observar de la Figura 1, todo árbol clasificador inicia con un nodo al que pertenecen todos los casos de la muestra que se quiere clasificar llamado nodo raíz (en la Figura es el nodo negro). El resto de los nodos se dividen en nodos intermedios o no terminales (nodos en color blanco) y nodos hoja o nodos terminales, es decir, nodos que no se van a dividir más (nodos de color azul en la Figura 1) [23].

En la fase de construcción cada nodo hoja se hace corresponder con una categoría concreta de la variable clase; así, los nodos hoja representan las diferentes particiones en las que se ha dividido el espacio de clasificación. Los nodos que “cuelgan” de un nodo concreto se les llama nodos hijo de dicho nodo, y al nodo del que parten las flechas o ramas se le denomina nodo padre [23].

A la hora de clasificar cada patrón, el punto de partida es el nodo raíz y, dependiendo de los valores de la variable predictora por la que se pregunta, los casos se van distribuyendo por los nodos hijo. Este proceso se repite en cada nodo hasta llegar los nodos hoja. En el árbol de la Figura 1 a todos los casos de la muestra a clasificar se pregunta por el valor que tiene la variable continua X_2 . Los casos que tengan un valor menor o igual que w irán al nodo intermedio izquierdo, los que no cumplan con esta regla irán al nodo de la derecha. Estos últimos casos quedan clasificados con

la clase que en la fase de entrenamiento se le ha asociado a dicho nodo hoja. A los nodos que les ha correspondido el nodo intermedio se les vuelve a hacer otra pregunta. Como consecuencia de esta pregunta, algunos quedaran clasificados en otro nodo hoja y el resto a nodos no terminales y así sucesivamente [23].

8.3 Redes Bayesianas

Las redes bayesianas modelan un fenómeno mediante un conjunto de variables y las relaciones de dependencia entre ellas. Estos modelos permiten hacer inferencia bayesiana; es decir, estimar la probabilidad posterior de las variables no conocidas, en función de las variables conocidas. Estas redes son una representación gráfica de dependencias para razonamiento probabilístico, en la cual los nodos representan variables aleatorias y los arcos representan relaciones de dependencia directa entre las variables [24].

En la Figura 2 se muestra una representación gráfica de una red bayesiana hipotética de cierto conocimiento sobre medicina.

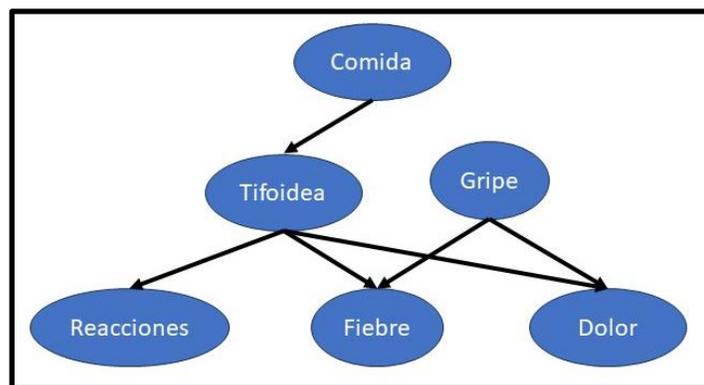


Figura 2. Ejemplo de red bayesiana.

Referencia [24]

En el ejemplo mostrado en la Figura 2, los nodos representan enfermedades, síntomas y factores que causan algunas enfermedades. La variable a la que apunta un arco es dependiente de la que esta en el origen de éste. La topología o estructura de la red proporciona información sobre las dependencias probabilísticas entre las variables. Esta red también representa las independencias condicionales de un variable o un conjunto de variables dadas otra u otras variables [24].

En una red bayesiana todas las relaciones de independencia condicional representadas en el grafo corresponden a relaciones de independencia en la distribución de probabilidad; dichas independencias simplifican la representación del conocimiento y el razonamiento, para el caso de X independiente de Y dado Z [24]:

- Independencia en la distribución $P(X|Y, Z) = P(X|Z)$.
- Independencia en el grafo $I\langle X|Z|Y\rangle$.

La probabilidad conjunta se especifica por el producto de las probabilidades de cada variable dados sus padres [24]:

$$P(x_1, x_2, x_3, \dots, x_n) = \prod_{i=1}^n P(x_i|P_a(x_i)) \quad (1)$$

Un clasificador bayesiano obtiene la probabilidad posterior de cada clase C_i , usando la regla de Bayes, como el producto de la probabilidad *a priori* de la clase por la probabilidad condicional de los atributos (E) dada la clase, dividido por la probabilidad de los atributos [24]

$$P(C_i|E) = \frac{P(C_i)P(E|C_i)}{P(E)} \quad (2)$$

Un clasificador bayesiano simple (Naive Bayes classifier, NBC) asume que los atributos son independientes entre sí dada la clase, así que la probabilidad se puede obtener por el producto de las probabilidades condicionales individuales de cada atributo dado el nodo clase [24]:

$$P(C_i|E) = \frac{P(C_i)P(E_1|C_i)P(E_2|C_i)P(E_3|C_i)\dots P(E_n|C_i)}{P(E)} \quad (3)$$

Donde n es el número de atributos.

Dado que los atributos son condicionalmente independientes dada la clase, no existen arcos entre ellos. La Figura 3 muestra una representación de un clasificador bayesiano simple con n número de atributos.

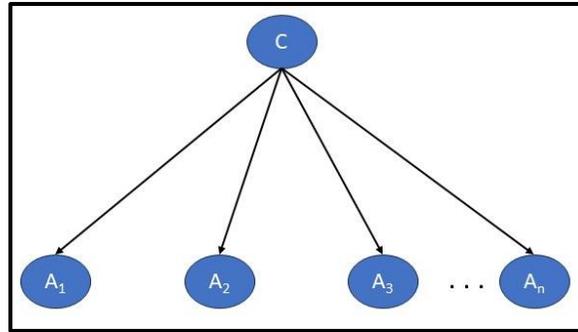


Figura 3. Clasificador bayesiano simple.

Referencia [24]

Como la estructura de un clasificador bayesiano simple ya está definido, únicamente es necesario aprender los parámetros asociados, que son [24]:

$P(C)$: vector de probabilidades *a priori* para cada clase

$P(E_i | C)$: matriz de probabilidad condicional para cada atributo de la clase.

Los parámetros se pueden estimar fácilmente a partir de los datos, en base a las frecuencias. Además, el denominador de la ecuación (3) no es necesario, ya que es una constante, es decir, no depende de la clase. Al final se puede simplemente normalizar las probabilidades posteriores de cada clase haciendo que sumen uno.

Como el clasificador bayesiano simple asume que los atributos son independientes dada la clase, si esto se llegara a cumplir, existen dos alternativas básicas. Transforma la estructura del clasificador a una red bayesiana, introduciendo arcos dirigidos entre los atributos dependientes; la desventaja es que se pierde la simplicidad ya que aprender el modelo y después clasificar nuevos casos llega a ser mas complejo. La otra alternativa es transformar la estructura, pero mantener una estructura de árbol [24].

8.3.1 Aprendizaje de redes bayesianas

El aprendizaje, en general, de redes bayesianas consiste en inducir un modelo, estructura y parámetros asociados, a partir de datos, este aprendizaje puede dividirse naturalmente en dos partes [24]:

1. Aprendizaje paramétrico. Dada la estructura, obtener las probabilidades asociadas.
2. Aprendizaje estructural. Obtener la estructura o topología de la red.

8.3.1.1 Aprendizaje paramétrico

Cuando se tienen datos completos y suficientes para todas las variables en el modelo, es relativamente fácil obtener los parámetros, asumiendo que la estructura está dada. El método más común es el llamado *estimador de máxima verosimilitud* bajo el cual se estiman las probabilidades en base a las frecuencias de los datos [24].

En la práctica, en muchas ocasiones los datos no están completos; los dos tipos básicos de información incompleta son:

Valores faltantes: Faltan algunos valores de algunas de las variables en algunos casos.

Nodos ocultos: Faltan todos los valores de una variable.

8.3.1.2 Aprendizaje estructural

El aprendizaje estructural consiste en encontrar las relaciones de dependencia entre las variables, de forma que se pueda determinar la topología o la estructura de la red bayesiana.

El *aprendizaje de árboles* es uno de los métodos de aprendizaje estructural; el cual se basa en el algoritmo desarrollado por Chow y Liu para aproximar una distribución de probabilidad por un producto de probabilidades de segundo orden [24].

La especificación de los parámetros de la red Naive Bayes consiste en estimar los parámetros de las distribuciones de probabilidad que se definen en la red. Esto es, definida la red Naive Bayes se tendrá que estimar los parámetros de la distribución de probabilidad de la variable clase $P(C)$ y de las distribuciones condicionales de los atributos dada la clase $P(A_i / C)$, tal como se muestra en la Figura 4[5].

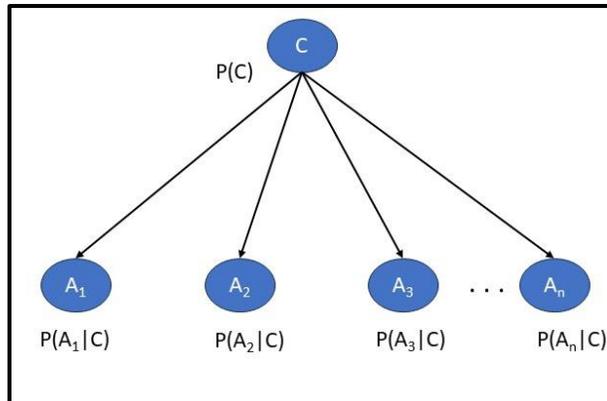


Figura 4. Red del clasificador Naive.

Referencia [5]

La estimación de los parámetros dependerá del tipo de variable o atributo A_i , que puede ser discreto o continuo.

Para atributos discreto, la distribución de probabilidad de la clase $P(C)$ y de la distribución condicional $P(A_i | C)$ se basa en la frecuencia relativa de ocurrencia de los sucesos correspondientes, que determinan el conjunto de datos. El valor estimado de probabilidad será:

$$P(a_i | C = c) = \frac{n(a_i \cap C = c)}{n(C = c)} \quad (4)$$

Donde

$n(a_i \cap C = c)$: es el número de casos del conjunto de datos en que la variable A_i toma el valor a_i y su padre (clase) C toma el valor de c .

$n(C = c)$: es el número de casos del conjunto de datos en que la clase C toma el valor de c .

Aun cuando esta estimación es una buena aproximación, en algunos casos suele suceder que los casos del conjunto de datos no contemplan todas las combinaciones de valores de la variable clase con las variables predictorias, con lo que se llega a una estimación de parámetros con abundancia de ceros [5].

El estimador basado en *ley de sucesión de Laplace* [25], en la cual en lugar de usar directamente la definición clásica de probabilidad ($P(A) = \frac{\text{Casos favorables}}{\text{Total de espacio muestral}}$) se utiliza como estimación el número que se obtiene del cociente

$$\frac{n(a_i \cap C=c)+1}{n(C=c)+r_{A_i}} \quad (5)$$

Es decir, el número de casos favorables más uno, dividido por el número de casos totales más el número de valores posibles del atributo A_i denotado por r_{A_i} [5].

En el caso de atributos continuos, el clasificador Naive bayes puede ser aplicado de dos formas alternativas [5]:

1. Aplicar previamente un método de discretización de la variable continua.
2. Asumiendo una distribución para cada variable predictora, que por lo general se considera gaussiana con media y varianza estimada de los datos.

Para este segundo caso, el clasificador Naive bayes supone que la variable predictora sigue una distribución normal, por lo que se requiere el cálculo, con los datos conocidos, de la media μ y la desviación estándar σ condicionadas a cada valor de la variable clase [5]

$$P(A_i|c) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)} \quad (6)$$

Teniendo el inconveniente de que los datos no siempre siguen una distribución normal.

9 Desarrollo

Debido al cambio constante de personal académico, se requiere una herramienta que permita dar claridad en la decisión sobre la contratación del nuevo personal académico que ayude en la asignación de la categoría correspondiente de acuerdo con las características curriculares del nuevo personal.

Para el desarrollo del modelo y por cuestiones de manejo de información, se desarrolla un cambio en el recuento de la base de datos del personal académico del TESCO. Se considera una plantilla de 232 trabajadores bajo la dirección académica; la categoría para la clasificación del personal académico es:

- ✓ Profesor de Asignatura A
- ✓ Profesor de asignatura B
- ✓ Profesor de tiempo completo asociado A
- ✓ Profesor de tiempo completo asociado B
- ✓ Profesor de tiempo completo asociado C
- ✓ Profesor de tiempo completo Titular A

En la Tabla 1, se muestra la distribución del personal en cada una de las categorías.

Tabla 1. Personal por categoría del personal académico.

Categoría	Porcentaje	Personal
Asignatura A	10%	23
Asignatura B	20%	46
Asociado A	30%	70
Asociado B	20%	46
Asociado C	15%	35
Titular A	5%	12

En la Tabla 2 se muestran los atributos para cada categoría.

Tabla 2. Características curriculares del personal académico.

Categoría	Características Curriculares
Experiencia profesional	Menos de 5 años Entre 5 y 10 años Mas de 10 años
Experiencia laboral	Menos de 5 años Entre 5 y 10 años Mas de 10 años
Posgrado	Si No
Perfil deseable	Si No
SNI	Si No

En la Figura 5 se muestra la estructura para el clasificador del problema planteado.

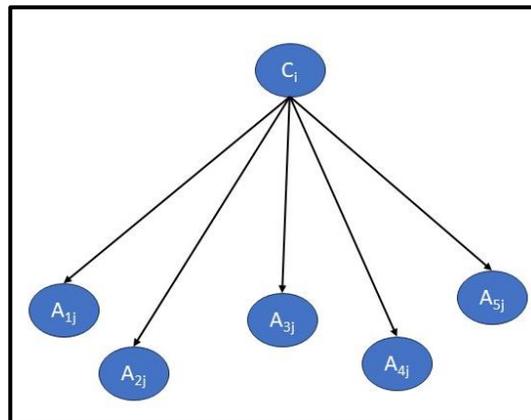


Figura 5. Red del clasificador Naive.

Donde

C_i = Categorías

$i = 1$, Asignatura A

$i = 2$, Asignatura B

$i = 3$, Asociado A

$i = 4$, Asociado B

$i = 5$, Asociado C

$i = 6$, Titular A

A_{1j} = Experiencia Docente

$j = 1$, Menos de 5 años

$j = 2$, Entre 5 y 10 años

$j = 3$, Mas de 10 años

A_{2j} = Experiencia Profesional

$j = 1$, Menos de 5 años

$j = 2$, Entre 5 y 10 años

$j = 3$, Mas de 10 años

A_{3j} = Posgrado

$j = 1$, Si

$j = 2$, No

A_{4j} = Perfil deseable

$j = 1$, Si

$j = 2$, No

A_{5j} = SNI

$j = 1$, Si

$j = 2$, No

En la Figura 6 se muestra un fragmento de los datos recopilados para el desarrollo del modelo. Se puede observar en la figura que los atributos que contienen los profesores se marcan como 1, y los que no los tienen se dejan vacíos.

Profeso	Plaza	EXPERIENCIA DOCENTE			EXPERIENCIA PROFESIONAL			Posgrado	Perfil	SNI
		Menos de 5 año	Entre 5 y 10 año	Mas de 10 año	Menos de 5 años	Entre 5 y 10 años	Mas de 10 años			
1	Asignatura A	1					1			
2	Asignatura A	1					1			
3	Asignatura A	1					1			
4	Asignatura A	1				1				
5	Asignatura A	1				1				
6	Asignatura A	1				1				
7	Asignatura A	1				1				
8	Asignatura A		1			1				
18	Asignatura B	1				1				
19	Asignatura B	1				1				
20	Asignatura B	1				1				
21	Asignatura B	1				1				
22	Asignatura B	1				1				
23	Asignatura B	1				1				
24	Asignatura B	1				1				
25	Asignatura B	1				1				
26	Asignatura B	1				1				
27	Asignatura B	1				1				
28	Asignatura B	1				1				
29	Asignatura B	1				1				
30	Asignatura B	1				1				
31	Asignatura B	1				1				
61	Asociado A	1				1		1	1	
62	Asociado A	1				1		1	1	
63	Asociado A	1				1		1	1	
64	Asociado A	1				1		1	1	
65	Asociado A	1				1		1	1	
66	Asociado A	1				1		1	1	
67	Asociado A	1				1		1	1	
68	Asociado A	1				1		1	1	
69	Asociado A	1				1		1	1	
70	Asociado A	1				1		1	1	
71	Asociado A	1				1		1	1	
72	Asociado A		1			1		1	1	
73	Asociado A		1			1		1	1	
74	Asociado A		1			1		1	1	
75	Asociado A		1			1		1	1	
76	Asociado A		1			1		1	1	
77	Asociado A		1			1		1	1	
78	Asociado A		1			1		1	1	
79	Asociado A		1			1		1	1	
80	Asociado A		1			1		1	1	
81	Asociado A		1			1		1	1	
107	Asociado B		1			1		1	1	
108	Asociado B		1			1		1	1	
109	Asociado B		1			1		1	1	
110	Asociado B		1			1		1	1	
111	Asociado B		1			1		1	1	
112	Asociado B			1		1				
113	Asociado B			1		1				
114	Asociado B			1		1				
115	Asociado B			1		1				
128	Asociado C	1				1		1	1	
129	Asociado C	1				1		1	1	
130	Asociado C	1				1		1	1	
131	Asociado C	1				1		1	1	
132	Asociado C	1				1		1	1	
133	Asociado C	1				1		1	1	
134	Asociado C	1				1		1	1	
170	Titular A	1					1	1	1	1
171	Titular A	1					1	1	1	1
172	Titular A	1				1		1	1	1
173	Titular A	1				1		1	1	1
175	Titular A		1			1		1	1	
176	Titular A		1			1		1	1	
177	Titular A			1		1		1	1	

Figura 6. Datos recopilados.

Se seleccionan aleatoriamente 180 registros de la plantilla adscrita a la dirección académica para desarrollar las distribuciones de probabilidad, los restantes 52 se utilizan para probar el modelo desarrollado.

En la Figura 7 se muestran una muestra de estos 180 registros.

Profeso	Plaza	EXPERIENCIA DOCENTE			EXPERIENCIA PROFESIONAL			Posgrado	Perfil	SNI
		Menos de 5 año	Entre 5 y 10 año	Mas de 10 año	Menos de 5 años	Entre 5 y 10 años	Mas de 10 años			
1	Asignatura A	1					1			
2	Asignatura A	1					1			
3	Asignatura A	1					1			
4	Asignatura A	1				1				
5	Asignatura A	1				1				
6	Asignatura A	1				1				
7	Asignatura A	1				1				
8	Asignatura A		1			1				
18	Asignatura B	1				1				
19	Asignatura B	1				1				
20	Asignatura B	1				1				
21	Asignatura B	1				1				
22	Asignatura B	1				1				
23	Asignatura B	1				1				
24	Asignatura B	1				1				
25	Asignatura B	1				1				
26	Asignatura B	1				1				
27	Asignatura B	1				1				
28	Asignatura B	1				1				
29	Asignatura B	1			1					
30	Asignatura B	1			1					
31	Asignatura B	1			1					
61	Asociado A	1				1		1	1	
62	Asociado A	1				1		1	1	
63	Asociado A	1				1		1	1	
64	Asociado A	1				1		1	1	
65	Asociado A	1				1		1	1	
66	Asociado A	1				1		1	1	
67	Asociado A	1				1		1	1	
68	Asociado A	1				1		1	1	
69	Asociado A	1				1		1	1	
70	Asociado A	1			1			1	1	
71	Asociado A	1			1			1	1	
72	Asociado A		1		1			1	1	
73	Asociado A		1		1			1	1	
74	Asociado A		1		1			1	1	
75	Asociado A		1		1			1	1	
76	Asociado A		1		1			1	1	
77	Asociado A		1		1			1	1	
78	Asociado A		1		1			1	1	
79	Asociado A		1		1			1	1	
80	Asociado A		1		1			1	1	
81	Asociado A		1		1			1	1	
107	Asociado B		1			1		1	1	
108	Asociado B		1		1			1	1	
109	Asociado B		1		1			1	1	
110	Asociado B		1		1			1	1	
111	Asociado B		1		1			1	1	
112	Asociado B			1	1					
113	Asociado B			1	1					
114	Asociado B			1	1					
115	Asociado B			1	1					
128	Asociado C	1				1		1	1	
129	Asociado C	1				1		1	1	
130	Asociado C	1				1		1	1	
131	Asociado C	1				1		1	1	
132	Asociado C	1				1		1	1	
133	Asociado C	1				1		1	1	
134	Asociado C	1				1		1	1	
170	Titular A	1					1	1	1	1
171	Titular A	1					1	1	1	1
172	Titular A	1				1		1	1	1
173	Titular A	1				1		1	1	1
175	Titular A		1		1			1	1	
176	Titular A		1		1			1	1	
177	Titular A			1	1			1		
178	Titular A			1	1					
179	Titular A			1	1					
180	Titular A			1	1					

Figura 7. Muestra aleatoria.

En la Tabla 3 se muestra los resultados referentes a las características curriculares observadas en la muestra aleatoria.

Tabla 3. Características curriculares muestra aleatoria.

Experiencia Docente	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
menos de 5 años	7	19	23	5	29	4
entre 5 y 10 años	5	7	19	6	21	3
más de 10 años	1	9	10	4	4	4
Experiencia Profesional	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
menos de 5 años	5	20	31	8	28	6
entre 5 y 10 años	5	11	19	4	17	3
más de 10 años	3	4	2	3	9	2
Posgrado	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
Si	0	0	40	11	45	8
No	13	35	12	4	9	3
Perfil deseable	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
Si	0	0	44	9	43	7
No	13	35	8	6	11	4
SNI	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
Si	0	0	0	0	0	4
No	13	35	52	15	54	7

Considerando que la estimación de la distribución de probabilidad de la categoría $P(C_i)$ y de la distribución condicional $P(a_{ij} / C_i)$ se basa en la frecuencia relativa de ocurrencia de los correspondientes sucesos, el valor para las probabilidades estimadas se obtiene a partir de la ecuación (4).

De la Tabla 3, también se observa que algunas características curriculares no cuentan con elementos, pero no en todos los casos, por lo que se aplica el corrector de Laplace descrito en la ecuación (5), con lo cual se evita que se asigne una probabilidad nula para que no se elimine por completo la probabilidad final.

En la Tabla 4 se muestran los datos con el corrector de Laplace.

Tabla 4. Muestra aleatoria aplicado el corrector de Laplace.

		experiencia docente				
Experiencia Docente	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
menos de 5 años	8	20	24	26	30	5
entre 5 y 10 años	6	8	20	7	22	4
mas de 10 años	2	10	11	5	5	5
		experiencia profesional				
Experiencia Profesional	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
menos de 5 años	6	21	32	9	29	7
entre 5 y 10 años	6	12	20	5	18	4
mas de 10 años	4	5	3	4	10	3
		posgrado				
Posgrado	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
Si	1	1	41	12	46	9
No	14	36	13	5	10	4
		perfil deseable				
Perfil deseable	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
Si	1	1	45	10	44	8
No	14	36	9	7	12	5
		SNI				
SNI	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
Si	1	1	1	1	1	5
No	14	36	53	16	55	8

En la Tabla 5 se muestra la probabilidad conjunta de cada una de las categorías estudiadas con sus respectivas características asociadas.

Tabla 5. Probabilidades conjuntas por categoría.

P(A1 Ci)		P(Requisito Experiencia Docente)		
Categoría i		menos de 5 años	entre 5 y 10 años	más de 10 años
	Asignatura A		1/2	3/8
Asigantura B		10/19	4/19	5/19
Asociado A		24/55	4/11	1/5
Asociado B		1/3	7/18	5/18
Asociado C		10/19	22/57	5/57
Titular A		4/11	3/11	4/11

P(A2 Ci)		P(Requisito Experiencia Profesional)		
Categoría i		menos de 5 años	entre 5 y 10 años	más de 10 años
	Asignatura A		3/8	3/8
Asigantura B		21/38	6/19	5/38
Asociado A		32/55	4/11	3/55
Asociado B		1/2	5/18	2/9
Asociado C		29/57	6/19	10/57
Titular A		6/11	3/11	2/11

P(A3 Ci)		P(Posgrado)	
Categoría i		Si	No
	Asignatura A		1/8
Asigantura B		1/19	18/19
Asociado A		10/13	3/13
Asociado B		11/15	4/15
Asociado C		5/6	1/6
Titular A		8/11	3/11

P(A4 Ci)		P(Perfil Deseable)	
Categoría i		Si	No
	Asignatura A		1/8
Asigantura B		1/37	36/37
Asociado A		11/13	2/13
Asociado B		3/5	2/5
Asociado C		43/54	11/54
Titular A		7/11	4/11

P(A5 Ci)		P(SNI)	
Categoría i		Si	No
	Asignatura A		1/8
Asigantura B		1/19	18/19
Asociado A		2/55	53/55
Asociado B		1/9	8/9
Asociado C		2/57	55/57
Titular A		4/11	7/11

En la Tabla 6 se muestran las probabilidades individuales para cada categoría.

Tabla 6. Probabilidades individuales por categoría.

Categoría	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Titular A
Recuento	16	38	55	18	57	11
P(Categoría _i)	16/195	38/195	11/39	6/65	16/65	11/195

En la Tabla 7 se muestran las 72 combinaciones posibles para las características curriculares del personal académico del Tecnológico de Estudios Superiores de Coacalco.

Tabla 7. Combinaciones de características curriculares.

	Experiencia Docente	Experiencia Profesional	Posgrado	Perfil Deseable	SNI
1	$x < 5$	$x < 5$	Si	Si	Si
2	$5 < x < 10$	$x < 5$	Si	Si	Si
3	$x > 10$	$x < 5$	Si	Si	Si
4	$x < 5$	$5 < x < 10$	Si	Si	Si
5	$5 < x < 10$	$5 < x < 10$	Si	Si	Si
6	$x > 10$	$5 < x < 10$	Si	Si	Si
7	$x < 5$	$x > 10$	Si	Si	Si
8	$5 < x < 10$	$x > 10$	Si	Si	Si
9	$x > 10$	$x > 10$	Si	Si	Si
10	$x < 5$	$x < 5$	No	Si	Si
11	$5 < x < 10$	$x < 5$	No	Si	Si
12	$x > 10$	$x < 5$	No	Si	Si
13	$x < 5$	$5 < x < 10$	No	Si	Si
14	$5 < x < 10$	$5 < x < 10$	No	Si	Si
15	$x > 10$	$5 < x < 10$	No	Si	Si
16	$x < 5$	$x > 10$	No	Si	Si
17	$5 < x < 10$	$x > 10$	No	Si	Si
18	$x > 10$	$x > 10$	No	Si	Si
19	$x < 5$	$x < 5$	Si	No	Si
20	$5 < x < 10$	$x < 5$	Si	No	Si
21	$x > 10$	$x < 5$	Si	No	Si
22	$x < 5$	$5 < x < 10$	Si	No	Si
23	$5 < x < 10$	$5 < x < 10$	Si	No	Si
24	$x > 10$	$5 < x < 10$	Si	No	Si
25	$x < 5$	$x > 10$	Si	No	Si
26	$5 < x < 10$	$x > 10$	Si	No	Si
27	$x > 10$	$x > 10$	Si	No	Si
28	$x < 5$	$x < 5$	No	No	Si
29	$5 < x < 10$	$x < 5$	No	No	Si
30	$x > 10$	$x < 5$	No	No	Si
31	$x < 5$	$5 < x < 10$	No	No	Si
32	$5 < x < 10$	$5 < x < 10$	No	No	Si
33	$x > 10$	$5 < x < 10$	No	No	Si

Tabla 7 Continuación. Combinaciones de características curriculares.

	Experiencia Docente	Experiencia Profesional	Posgrado	Perfil Deseable	SNI
34	$x < 5$	$x > 10$	No	No	Si
35	$5 < x < 10$	$x > 10$	No	No	Si
36	$x > 10$	$x > 10$	No	No	Si
37	$x < 5$	$x < 5$	Si	Si	No
38	$5 < x < 10$	$x < 5$	Si	Si	No
39	$x > 10$	$x < 5$	Si	Si	No
40	$x < 5$	$5 < x < 10$	Si	Si	No
41	$5 < x < 10$	$5 < x < 10$	Si	Si	No
42	$x > 10$	$5 < x < 10$	Si	Si	No
43	$x < 5$	$x > 10$	Si	Si	No
44	$5 < x < 10$	$x > 10$	Si	Si	No
45	$x > 10$	$x > 10$	Si	Si	No
46	$x < 5$	$x < 5$	No	Si	No
47	$5 < x < 10$	$x < 5$	No	Si	No
48	$x > 10$	$x < 5$	No	Si	No
49	$x < 5$	$5 < x < 10$	No	Si	No
50	$5 < x < 10$	$5 < x < 10$	No	Si	No
51	$x > 10$	$5 < x < 10$	No	Si	No
52	$x < 5$	$x > 10$	No	Si	No
53	$5 < x < 10$	$x > 10$	No	Si	No
54	$x > 10$	$x > 10$	No	Si	No
55	$x < 5$	$x < 5$	Si	No	No
56	$5 < x < 10$	$x < 5$	Si	No	No
57	$x > 10$	$x < 5$	Si	No	No
58	$x < 5$	$5 < x < 10$	Si	No	No
59	$5 < x < 10$	$5 < x < 10$	Si	No	No
60	$x > 10$	$5 < x < 10$	Si	No	No
61	$x < 5$	$x > 10$	Si	No	No
62	$5 < x < 10$	$x > 10$	Si	No	No
63	$x > 10$	$x > 10$	Si	No	No
64	$x < 5$	$x < 5$	No	No	No
65	$5 < x < 10$	$x < 5$	No	No	No
66	$x > 10$	$x < 5$	No	No	No
67	$x < 5$	$5 < x < 10$	No	No	No
68	$5 < x < 10$	$5 < x < 10$	No	No	No
69	$x > 10$	$5 < x < 10$	No	No	No
70	$x < 5$	$x > 10$	No	No	No
71	$5 < x < 10$	$x > 10$	No	No	No
72	$x > 10$	$x > 10$	No	No	No

10 Resultados

Con los datos de la Tabla 5 y Tabla 6 se desarrolla el cálculo de probabilidades de predicción, para cada una de las 72 combinaciones posibles de categoría – característica, por medio de la ecuación

$$P(C_i|a_{1j}, a_{2j}, a_{3j}, a_{4j}, a_{5j}) = \alpha P(C_i)P(a_{1j}|C_i)P(a_{2j}|C_i)P(a_{3j}|C_i)P(a_{4j}|C_i)P(a_{5j}|C_i) \quad (7)$$

Donde α es la constancia de proporcionalidad.

A continuación, se muestran algunos de los cálculos desarrollados

Alternativa X<5, X<5, Si, Si, Si

Asignatura A

$$P(C_1|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha P(C_1)P(a_{11}|C_1)P(a_{21}|C_1)P(a_{31}|C_1)P(a_{41}|C_1)P(a_{51}|C_1)$$

$$P(C_1|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha \left(\frac{16}{195}\right) \left(\frac{1}{2}\right) \left(\frac{3}{8}\right) \left(\frac{1}{8}\right) \left(\frac{1}{8}\right) \left(\frac{1}{8}\right) = 0.00003 \alpha \quad (8)$$

Asignatura B

$$P(C_2|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha P(C_2)P(a_{11}|C_2)P(a_{21}|C_2)P(a_{31}|C_2)P(a_{41}|C_2)P(a_{51}|C_2)$$

$$P(C_2|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha \left(\frac{38}{195}\right) \left(\frac{10}{19}\right) \left(\frac{21}{38}\right) \left(\frac{1}{19}\right) \left(\frac{1}{37}\right) \left(\frac{1}{19}\right) = 0.0000042 \alpha \quad (9)$$

Asociado A

$$P(C_3|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha P(C_3)P(a_{11}|C_3)P(a_{21}|C_3)P(a_{31}|C_3)P(a_{41}|C_3)P(a_{51}|C_3)$$

$$P(C_3|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha \left(\frac{11}{39}\right) \left(\frac{24}{55}\right) \left(\frac{32}{55}\right) \left(\frac{10}{13}\right) \left(\frac{11}{13}\right) \left(\frac{2}{55}\right) = 0.0000042 \alpha \quad (10)$$

Asociado B

$$P(C_4|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha P(C_4)P(a_{11}|C_4)P(a_{21}|C_4)P(a_{31}|C_4)P(a_{41}|C_4)P(a_{51}|C_4)$$

$$P(C_4|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha \left(\frac{6}{65}\right) \left(\frac{1}{3}\right) \left(\frac{1}{2}\right) \left(\frac{11}{15}\right) \left(\frac{3}{5}\right) \left(\frac{1}{9}\right) = 0.0007521 \alpha \quad (11)$$

Asociado C

$$P(C_5|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha P(C_5)P(a_{11}|C_5)P(a_{21}|C_5)P(a_{31}|C_5)P(a_{41}|C_5)P(a_{51}|C_5)$$

$$P(C_5|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha \left(\frac{19}{65}\right) \left(\frac{10}{19}\right) \left(\frac{29}{57}\right) \left(\frac{5}{6}\right) \left(\frac{43}{54}\right) \left(\frac{2}{57}\right) = 0.0015347 \alpha \quad (12)$$

Titular A

$$P(C_6|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha P(C_6)P(a_{11}|C_6)P(a_{21}|C_6)P(a_{31}|C_6)P(a_{41}|C_6)P(a_{51}|C_6)$$

$$P(C_6|a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) = \alpha \left(\frac{11}{195}\right) \left(\frac{4}{11}\right) \left(\frac{6}{11}\right) \left(\frac{8}{11}\right) \left(\frac{7}{11}\right) \left(\frac{4}{11}\right) = 0.0018830 \alpha \quad (13)$$

De las ecuaciones (8) a la (13) se tiene que cumplir

$$0.00003 \alpha + 0.0000042 \alpha + 0.0007521 \alpha + 0.0007521 \alpha + 0.0015347 \alpha + 0.0018830 \alpha = 1$$

De donde

$$0.0058990 \alpha = 1 \quad (14)$$

Por lo tanto

$$\alpha = \frac{1}{0.0058990} \approx 169.5196396 \quad (15)$$

En la Tabla 8 se muestran las probabilidades para la alternativa X<5, X<5, Si, Si, Si. Las cuales se obtienen sustituyendo el valor de α , encontrado en la ecuación (15), en las ecuaciones (8) a (13).

Tabla 8. Probabilidad por alternativa 1.

Alternativa X<5, X<5, Si, Si, Si	
$P(C_1 a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) =$	0.0050937
$P(C_2 a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) =$	0.0007194
$P(C_3 a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) =$	0.2873143
$P(C_4 a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) =$	0.1275020
$P(C_5 a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) =$	0.2601625
$P(C_6 a_{11}, a_{21}, a_{31}, a_{41}, a_{51}) =$	0.3192081

El clasificador bayesiano asigna a un individuo a la clase con mayor probabilidad a posteriori, por ello, para un aspirante académico con las características siguientes experiencia docente menor de 5 años, experiencia profesional menor de 5 años, con posgrado, con perfil deseable y miembro del SNI, será asignado a la categoría Titula A, ya que se tiene una mayor probabilidad (0.3192081).

Cálculo

A continuación, se muestran algunos de los cálculos desarrollados

Alternativa X<5, X<5, No, No, No

Asignatura A

$$\begin{aligned} P(C_1|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha P(C_1)P(a_{11}|C_1)P(a_{21}|C_1)P(a_{32}|C_1)P(a_{42}|C_1)P(a_{52}|C_1) \\ P(C_1|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha \left(\frac{16}{195}\right) \left(\frac{1}{2}\right) \left(\frac{3}{8}\right) \left(\frac{7}{8}\right) \left(\frac{7}{8}\right) \left(\frac{7}{8}\right) = 0.0103065 \alpha \end{aligned} \quad (16)$$

Asignatura B

$$\begin{aligned} P(C_2|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha P(C_2)P(a_{11}|C_2)P(a_{21}|C_2)P(a_{32}|C_2)P(a_{42}|C_2)P(a_{52}|C_2) \\ P(C_2|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha \left(\frac{38}{195}\right) \left(\frac{10}{19}\right) \left(\frac{21}{38}\right) \left(\frac{18}{19}\right) \left(\frac{36}{37}\right) \left(\frac{18}{19}\right) = 0.0208404 \alpha \end{aligned} \quad (17)$$

Asociado A

$$\begin{aligned} P(C_3|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha P(C_3)P(a_{11}|C_3)P(a_{21}|C_3)P(a_{32}|C_3)P(a_{42}|C_3)P(a_{52}|C_3) \\ P(C_3|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha \left(\frac{11}{39}\right) \left(\frac{24}{55}\right) \left(\frac{32}{55}\right) \left(\frac{3}{13}\right) \left(\frac{2}{13}\right) \left(\frac{53}{55}\right) = 0.0007127 \alpha \end{aligned} \quad (18)$$

Asociado B

$$\begin{aligned} P(C_4|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha P(C_4)P(a_{11}|C_4)P(a_{21}|C_4)P(a_{32}|C_4)P(a_{42}|C_4)P(a_{52}|C_4) \\ P(C_4|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha \left(\frac{6}{65}\right) \left(\frac{1}{3}\right) \left(\frac{1}{2}\right) \left(\frac{4}{15}\right) \left(\frac{2}{5}\right) \left(\frac{8}{9}\right) = 0.0012966 \alpha \end{aligned} \quad (19)$$

Asociado C

$$\begin{aligned} P(C_5|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha P(C_5)P(a_{11}|C_5)P(a_{21}|C_5)P(a_{32}|C_5)P(a_{42}|C_5)P(a_{52}|C_5) \\ P(C_5|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha \left(\frac{19}{65}\right) \left(\frac{10}{19}\right) \left(\frac{29}{57}\right) \left(\frac{1}{6}\right) \left(\frac{11}{54}\right) \left(\frac{55}{57}\right) = 0.0007198 \alpha \end{aligned} \quad (20)$$

Titular A

$$\begin{aligned} P(C_6|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha P(C_6)P(a_{11}|C_6)P(a_{21}|C_6)P(a_{32}|C_6)P(a_{42}|C_6)P(a_{52}|C_6) \\ P(C_6|a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) &= \alpha \left(\frac{11}{195}\right) \left(\frac{4}{11}\right) \left(\frac{6}{11}\right) \left(\frac{3}{11}\right) \left(\frac{4}{11}\right) \left(\frac{7}{11}\right) = 0.0010271 \alpha \end{aligned} \quad (21)$$

De las ecuaciones (16) a la (21) se tiene que cumplir

$$0.0103065 \alpha + 0.0208404 \alpha + 0.0007127 \alpha + 0.0012966 \alpha + 0.0007198 \alpha + 0.0010271 \alpha = 1$$

De donde

$$0.0349031 \alpha = 1 \quad (22)$$

Por lo tanto

$$\alpha = \frac{1}{0.0349031} \approx 28.6507876 \quad (23)$$

En la Tabla 9 se muestran las probabilidades para la alternativa X<5, X<5, No, No, No. Las cuales se obtienen sustituyendo el valor de α , encontrado en la ecuación (23), en las ecuaciones (16) a (21).

Tabla 8. Probabilidad por alternativa 64.

Alternativa X<5, X<5, Si, Si, Si	
$P(C_1 a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) =$	0.2952891
$P(C_2 a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) =$	0.5970939
$P(C_3 a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) =$	0.0204190
$P(C_4 a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) =$	0.0371490
$P(C_5 a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) =$	0.0206218
$P(C_6 a_{11}, a_{21}, a_{32}, a_{42}, a_{52}) =$	0.0294272

Por lo tanto un aspirante académico con las características siguientes experiencia docente menor de 5 años, experiencia profesional menor de 5 años, sin posgrado, sin perfil deseable y no miembro del SNI, será asignado a la categoría Asignatura B, ya que se tiene una mayor probabilidad (0.5970939).

Este proceso se continua para obtener las probabilidades de las 72 alternativas. El resultado se muestra en la Tabla 9.

Tabla 9. Probabilidad 72 alternativas.

	Experiencia Docente	Experiencia Profesional	Posgrado	Perfil Deseable	SNI	Categoría	Probabilidad
1	$x < 5$	$x < 5$	Si	Si	Si	Titular A	0.31920812
2	$5 < x < 10$	$x < 5$	Si	Si	Si	Asociado A	0.29110535
3	$x > 10$	$x < 5$	Si	Si	Si	Asociado A	0.99597308
4	$x < 5$	$5 < x < 10$	Si	Si	Si	Titular A	0.57117140
5	$5 < x < 10$	$5 < x < 10$	Si	Si	Si	Titular A	0.52057107
6	$x > 10$	$5 < x < 10$	Si	Si	Si	Titular A	0.72798625
7	$x < 5$	$x > 10$	Si	Si	Si	Titular A	0.62825222
8	$5 < x < 10$	$x > 10$	Si	Si	Si	Titular A	0.56713413
9	$x > 10$	$x > 10$	Si	Si	Si	Titular A	0.75055591
10	$x < 5$	$x < 5$	No	Si	Si	Titular A	0.58260539
11	$5 < x < 10$	$x < 5$	No	Si	Si	Titular A	0.54138826
12	$x > 10$	$x < 5$	No	Si	Si	Titular A	0.74254928
13	$x < 5$	$5 < x < 10$	No	Si	Si	Titular A	0.50773482
14	$5 < x < 10$	$5 < x < 10$	No	Si	Si	Titular A	0.46901332
15	$x > 10$	$5 < x < 10$	No	Si	Si	Titular A	0.69320677
16	$x < 5$	$x > 10$	No	Si	Si	Titular A	0.40926668
17	$5 < x < 10$	$x > 10$	No	Si	Si	Titular A	0.48168441
18	$x > 10$	$x > 10$	No	Si	Si	Titular A	0.70805676
19	$x < 5$	$x < 5$	Si	No	Si	Titular A	0.62454619
20	$5 < x < 10$	$x < 5$	Si	No	Si	Titular A	0.57316410
21	$x > 10$	$x < 5$	Si	No	Si	Titular A	0.75094699
22	$x < 5$	$5 < x < 10$	Si	No	Si	Titular A	0.55322897
23	$5 < x < 10$	$5 < x < 10$	Si	No	Si	Titular A	0.50836515
24	$x > 10$	$5 < x < 10$	Si	No	Si	Titular A	0.71200688
25	$x < 5$	$x > 10$	Si	No	Si	Titular A	0.56165475
26	$5 < x < 10$	$x > 10$	Si	No	Si	Titular A	0.65278679
27	$x > 10$	$x > 10$	Si	No	Si	Titular A	0.70417781
28	$x < 5$	$x < 5$	No	No	Si	Asignatura A	0.42898141
29	$5 < x < 10$	$x < 5$	No	No	Si	Asignatura A	0.49335797
30	$x > 10$	$x < 5$	No	No	Si	Titular A	0.34818147
31	$x < 5$	$5 < x < 10$	No	No	Si	Asignatura A	0.57240649
32	$5 < x < 10$	$5 < x < 10$	No	No	Si	Asignatura A	0.50299072
33	$x > 10$	$5 < x < 10$	No	No	Si	Asignatura A	0.33967219
34	$x < 5$	$x > 10$	No	No	Si	Asignatura A	0.63885374
35	$5 < x < 10$	$x > 10$	No	No	Si	Asignatura A	0.64191032
36	$x > 10$	$x > 10$	No	No	Si	Asignatura A	0.38249835

Tabla 9 Continuación. Probabilidad 72 alternativas.

37	$x < 5$	$x < 5$	Si	Si	No	Asociado C	0.37496788
38	$5 < x < 10$	$x < 5$	Si	Si	No	Asociado A	0.38408808
39	$x > 10$	$x < 5$	Si	Si	No	Asociado A	0.33925357
40	$x < 5$	$5 < x < 10$	Si	Si	No	Asociado C	0.39131161
41	$5 < x < 10$	$5 < x < 10$	Si	Si	No	Asociado A	0.36512823
42	$x > 10$	$5 < x < 10$	Si	Si	No	Asociado A	0.36938363
43	$x < 5$	$x > 10$	Si	Si	No	Asociado C	0.47565533
44	$5 < x < 10$	$x > 10$	Si	Si	No	Asociado C	0.41088394
45	$x > 10$	$x > 10$	Si	Si	No	Asociado B	0.39718253
46	$x < 5$	$x < 5$	No	Si	No	Asociado A	0.31290639
47	$5 < x < 10$	$x < 5$	No	Si	No	Asociado A	0.31766550
48	$x > 10$	$x < 5$	No	Si	No	Titular A	0.28345027
49	$x < 5$	$5 < x < 10$	No	Si	No	Asociado A	0.30705304
50	$5 < x < 10$	$5 < x < 10$	No	Si	No	Asociado B	0.28121097
51	$x > 10$	$5 < x < 10$	No	Si	No	Asociado A	0.29970375
52	$x < 5$	$x > 10$	No	Si	No	Asignatura A	0.25035922
53	$5 < x < 10$	$x > 10$	No	Si	No	Asociado B	0.30357503
54	$x > 10$	$x > 10$	No	Si	No	Asociado B	0.36678676
55	$x < 5$	$x < 5$	Si	No	No	Asociado C	0.24138236
56	$5 < x < 10$	$x < 5$	Si	No	No	Asociado B	0.34405420
57	$x > 10$	$x < 5$	Si	No	No	Asociado B	0.35602923
58	$x < 5$	$5 < x < 10$	Si	No	No	Asociado C	0.24272327
59	$5 < x < 10$	$5 < x < 10$	Si	No	No	Asociado B	0.30479343
60	$x > 10$	$5 < x < 10$	Si	No	No	Asociado B	0.34593452
61	$x < 5$	$x > 10$	Si	No	No	Asociado B	0.30367022
62	$5 < x < 10$	$x > 10$	Si	No	No	Asociado B	0.41309211
63	$x > 10$	$x > 10$	Si	No	No	Asociado B	0.45138220
64	$x < 5$	$x < 5$	No	No	No	Asignatura B	0.59709388
65	$5 < x < 10$	$x < 5$	No	No	No	Asignatura B	0.42813648
66	$x > 10$	$x < 5$	No	No	No	Asignatura B	0.67006444
67	$x < 5$	$5 < x < 10$	No	No	No	Asignatura B	0.48924135
68	$5 < x < 10$	$5 < x < 10$	No	No	No	Asignatura A	0.53613531
69	$x > 10$	$5 < x < 10$	No	No	No	Asignatura A	0.39947206
70	$x < 5$	$x > 10$	No	No	No	Asignatura A	0.52584236
71	$5 < x < 10$	$x > 10$	No	No	No	Asignatura A	0.62051213
72	$x > 10$	$x > 10$	No	No	No	Asignatura B	0.44916888

En la Figura 8 se muestra la prueba del modelo sobre los 52 registros faltantes.

Profesor	Plaza	EXPERIENCIA DOCENTE			EXPERIENCIA PROFESIONAL			Posgrado	Perfil	SNI	Columna1	Columna2
		Menos de 5 años	Entre 5 y 10 años	Mas de 10 años	Menos de 5 años	Entre 5 y 10 años	Mas de 10 años					
1	Asignatura A	1			1						11222	Incorrecto
2	Asignatura A	1			1						11222	Incorrecto
3	Asignatura A	1			1						11222	Incorrecto
4	Asignatura A		1		1						21222	Incorrecto
5	Asignatura A		1		1						21222	Incorrecto
6	Asignatura A		1			1					22222	Correcto
7	Asignatura A		1			1					22222	Correcto
8	Asignatura A		1			1					22222	Correcto
9	Asignatura A			1			1				33222	Incorrecto
10	Asignatura A			1			1				33222	Incorrecto
11	Asignatura B		1		1						21222	Correcto
12	Asignatura B		1			1					23222	Incorrecto
13	Asignatura B		1				1				23222	Incorrecto
14	Asignatura B		1				1				23222	Incorrecto
15	Asignatura B		1			1					21222	Correcto
16	Asignatura B			1	1						31222	Correcto
17	Asignatura B			1		1					32222	Incorrecto
18	Asignatura B			1			1				33222	Correcto
19	Asignatura B			1			1				33222	Correcto
20	Asignatura B			1			1				33222	Correcto
21	Asignatura B	1					1				13222	Incorrecto
22	Asociado A		1		1			1	1		21112	Correcto
23	Asociado A		1		1			1	1		21112	Correcto
24	Asociado A		1		1			1	1		21112	Correcto
25	Asociado A		1			1		1	1		23112	Incorrecto
26	Asociado A		1			1		1	1		23112	Incorrecto
27	Asociado A		1			1					23222	Correcto
28	Asociado A		1			1					23222	Correcto
29	Asociado A		1			1					23222	Correcto
30	Asociado A		1			1					23222	Correcto
31	Asociado A			1			1	1	1		33112	Incorrecto
32	Asociado A			1			1				33222	Incorrecto
33	Asociado A			1			1				33222	Incorrecto
34	Asociado A			1			1				33222	Incorrecto
35	Asociado A			1			1	1	1		33112	Incorrecto
36	Asociado A			1			1	1	1		33112	Incorrecto
37	Asociado A			1			1	1	1		33112	Incorrecto
38	Asociado A			1			1	1	1		33112	Incorrecto
39	Asociado A			1			1	1	1		33112	Incorrecto
40	Asociado B		1			1		1	1		22112	Incorrecto
41	Asociado B		1			1		1	1		22112	Incorrecto
42	Asociado B	1					1	1	1		13112	Incorrecto
43	Asociado B			1			1	1	1		33112	Correcto
44	Asociado B		1				1	1	1		23112	Incorrecto
45	Asociado B	1					1	1	1		13112	Incorrecto
46	Asociado B	1			1			1	1		11112	Incorrecto
47	Asociado B	1			1						11222	Correcto
48	Asociado B	1			1			1	1		11112	Incorrecto
49	Asociado B			1		1		1	1		32112	Incorrecto
50	Asociado B			1		1		1	1		32112	Incorrecto
51	Asociado B			1		1		1	1		32112	Incorrecto
52	Titular A			1		1		1	1		32112	Incorrecto

Figura 8. Evaluación del modelo.

Se puede observar que el modelo desarrollado presenta una certeza del 34%.

En la Figura 9 se observan los casos clasificados como incorrectos.

Profesor	Plaza	EXPERIENCIA DOCENTE			EXPERIENCIA PROFESIONAL			Posgrado	Perfil	SNI
		Menos de 5 años	Entre 5 y 10 años	Mas de 10 años	Menos de 5 años	Entre 5 y 10 años	Mas de 10 años			
4	Asignatura A		1		1					
5	Asignatura A		1		1					
9	Asignatura A			1			1			
10	Asignatura A			1			1			
1	Asignatura A	1			1					
2	Asignatura A	1			1					
3	Asignatura A	1			1					
21	Asignatura B	1					1			
12	Asignatura B		1			1				
13	Asignatura B		1				1			
14	Asignatura B		1				1			
17	Asignatura B			1		1				
25	Asociado A		1			1		1	1	
26	Asociado A		1			1		1	1	
31	Asociado A			1			1	1	1	
35	Asociado A			1			1	1	1	
36	Asociado A			1			1	1	1	
37	Asociado A			1			1	1	1	
38	Asociado A			1			1	1	1	
39	Asociado A			1			1	1	1	
32	Asociado A			1			1			
33	Asociado A			1			1			
34	Asociado A			1			1			
46	Asociado B	1			1			1	1	
48	Asociado B	1			1			1	1	
42	Asociado B	1					1	1	1	
45	Asociado B	1					1	1	1	
40	Asociado B		1			1		1	1	
41	Asociado B		1			1		1	1	
44	Asociado B		1				1	1	1	
49	Asociado B			1		1		1	1	
50	Asociado B			1		1		1	1	
51	Asociado B			1		1		1	1	
52	Titular A			1		1		1	1	

Figura 9. Evaluación incorrecta del modelo.

En la Figura 10 se observan los casos clasificados como correctos.

Profesor	Plaza	EXPERIENCIA DOCENTE			EXPERIENCIA PROFESIONAL			Posgrado	Perfil	SNI
		Menos de 5 años	Entre 5 y 10 años	Mas de 10 años	Menos de 5 años	Entre 5 y 10 años	Mas de 10 años			
6	Asignatura A		1			1				
7	Asignatura A		1			1				
8	Asignatura A		1			1				
11	Asignatura B		1		1					
15	Asignatura B		1			1				
16	Asignatura B			1	1					
18	Asignatura B			1			1			
19	Asignatura B			1			1			
20	Asignatura B			1			1			
22	Asociado A		1		1			1	1	
23	Asociado A		1		1			1	1	
24	Asociado A		1		1			1	1	
27	Asociado A		1			1				
28	Asociado A		1			1				
29	Asociado A		1			1				
30	Asociado A		1			1				
43	Asociado B			1			1	1	1	
47	Asociado B	1			1					

Figura 10. Evaluación correcta del modelo.

De las Figuras 9 y 10 se puede observar inconsistencia en los registros ya que a personal académico con las mismas características se les asigna categoría diferente.

11 Conclusiones y recomendaciones

Aun cuando el resultado del modelo presenta un resultado de 34% de asignaciones correctas, se hace la observación que los datos son modificados por cuestiones de privacidad de la información; sin embargo, el modelo desarrollado permite usar los datos originales y hacer el cálculo automáticamente.

El modelo es de gran ayuda al permitir dar claridad en la situación de decisión referente a la asignación de plazas de personal académico de nueva contratación.

También se hace notar que las inconsistencias en los datos se seguirán presentando ya que el Tecnológico de Estudios Superiores de Coacalco cuenta con plazas limitadas de cada categoría.

En el trabajo se consiguen los objetivos planteados referentes a la generación del modelo en base a las características curriculares de la planta docente y al cálculo de las probabilidades correspondientes.

Se recomienda que se actualice la base de datos conforme las características curriculares del personal académico cambien.

12 Bibliografía

- [1]Estado de México, “Gaceta de Gobierno ,” México, 1996. Accessed: Oct. 03, 2022. [Online]. Available: https://legislacion.edomex.gob.mx/ve_periodico_oficial
- [2]TECNOLÓGICO NACIONAL DE MÉXICO, “Lineamientos que establecen los requisitos y condiciones que debe reunir el personal académico que imparte planes y programas de estudio en los Institutos Tecnológicos Descentralizados,” <http://www.dgest.gob.mx/itd/academica-ditd>.
- [3]C. Malagón Luque, “Clasificadores bayesianos. El algoritmo Naïve Bayes,” 2003.
- [4]B. Sierra Araujo, *Aprendizaje automático: conceptos básicos y avanzados*, Primera., vol. 1. ESPAÑA: PEARSON - PRENTICE HALL, 2006.
- [5]F. De, C. Matemáticas, U. De Posgrado, A. Carlos, and A. Jaimes, “UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS OPTIMIZACIÓN DEL CLASIFICADOR ‘NAIVE BAYES’ USANDO ÁRBOL DE DECISIÓN C4.5 TESIS Para optar el Grado Académico de Magister en Estadística,” 2015.
- [6]D. W. Gerdes *et al.*, “A study on the application of decision tree algorithm in mobile marketing,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Oct. 2021. doi: 10.1088/1742-6596/2037/1/012033.
- [7]Z. A. Abutiheen, A. H. Aliwy, and K. B. S. Aljanabi, “Arabic text classification using master-slaves technique,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2018. doi: 10.1088/1742-6596/1032/1/012052.
- [8]M. Wasil, A. Sudianto, and Fathurrahman, “Application of the Decision Tree Method to Predict Student Achievement Viewed from Final Semester Values,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jul. 2020. doi: 10.1088/1742-6596/1539/1/012027.

- [9] R. E. Barrientos Martínez *et al.*, “ARTÍCULO ORIGINAL,” pp. 19–24, Sep. 2009, [Online]. Available: www.uv.mx/rm
- [10] D. W. Gerdes *et al.*, “ArborZ: Photometric redshifts using boosted decision trees,” *Astrophysical Journal*, vol. 715, no. 2, pp. 823–832, 2010, doi: 10.1088/0004-637X/715/2/823.
- [11] P. S. Broos, K. V. Getman, M. S. Povich, L. K. Townsley, E. D. Feigelson, and G. P. Garmire, “A naive bayes source classifier for X-ray sources,” *Astrophysical Journal, Supplement Series*, vol. 194, no. 1, May 2011, doi: 10.1088/0067-0049/194/1/4.
- [12] O. F. Althuwaynee, A. L. Balogun, A. Aydda, and T. Gumbo, “Classification of air pollutants API Inter-Correlation using decision tree algorithms,” in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Feb. 2020. doi: 10.1088/1755-1315/419/1/012022.
- [13] N. Gangsarwijaya, R. Anindhita, and D. L. Widaningrum, “Decision tree analysis approach to determine factors that affect the quote order lead time fulfillment,” in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1755-1315/426/1/012122.
- [14] Y. Kustiyahningsih, B. K. Khotimah, D. R. Anamisa, M. Yusuf, T. Rahayu, and J. Purnama, “Decision Tree C 4.5 Algorithm for Classification of Poor Family Scholarship Recipients,” *IOP Conf Ser Mater Sci Eng*, vol. 1125, no. 1, p. 012048, May 2021, doi: 10.1088/1757-899x/1125/1/012048.
- [15] A. Desiani, R. Primartha, M. Arhami, and O. Orsalan, “Naive Bayes classifier for infant weight prediction of hypertension mother,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Aug. 2019. doi: 10.1088/1742-6596/1282/1/012005.
- [16] N. R. Nurwulan and G. Selamaj, “Human daily activities recognition using decision tree,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1833/1/012039.
- [17] D. Nurcahyono, W. P. Putra, A. Najib, and T. R. Tulili, “Analysis sentiment in social media against election using the method naive Bayes,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1511/1/012003.

- [18] F. Syukmana *et al.*, “Predicting Relegation Clubs in Italian Serie A with Method based C4.5 Decision Tree Algorithm,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1742-6596/1471/1/012016.
- [19] P. A. Rahayuningsih, R. Maulana, W. Irmayani, D. Saputra, and D. Purwaningtias, “Feature Dependent Naïve Bayes for Network Intrusion Detection System,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012023.
- [20] Y. Wang and D. Duan, “Research on risk assessment of clients before loan based on decision tree algorithm,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1774/1/012056.
- [21] Samsir *et al.*, “Naives Bayes Algorithm for Twitter Sentiment Analysis,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jun. 2021. doi: 10.1088/1742-6596/1933/1/012019.
- [22] IBM, “¿Qué es el aprendizaje supervisado?”
- [23] J. Murgueza, Pérez. Jesús María, I. Gurrutuxaga, and J. I. Martín, “Árboles de Clasificación,” in *Aprendizaje automático*, vol. 1, España: Pearson - Educación, 2006, pp. 41–58.
- [24] L. E. Sucar, “Redes bayesianas,” in *Aprendizaje automático*, vol. 1, México: Pearson - Prentice Hall, 2006, pp. 77–100.
- [25] “The_MIT_Press_Probabilistic_Machine_Lear”.

Clasificador Naive como herramienta de decisión en la contratación de personal académico del TESCO.

Viveros-Rosas Leopoldo[#], Díaz-Téllez Rebeca^{*}; Pérez-Torres José Roberto[&];
Chew-Hernández Mario Luis[#] y Vega-González José Antonio[^];

[#]Subdirección de Estudios Profesionales “B” – TESCO, México.

^{*}División de Ingeniería en Gestión Empresarial – TESCO, México.

[^]División de Ingeniería Civil – TESCO, México.

[&] División de Ingeniería en Sistemas Computacionales – TESCO, México

Avenida 16 de Septiembre Número 54, Cabecera Municipal, Coacalco Estado de México C.P 55700

leopoldo@tesco.edu.mx, rebeca.sub.a@tesco.edu.mx, jose.roberto@tesco.edu.mx,

mario@tesco.edu.mx, a.vega@tesco.edu.mx.

Resumen - Se utiliza el modelo clasificador Naive como propuesta de herramienta para dar claridad en la decisión sobre las alternativas en la contratación de personal académico del Tecnológico de Estudios Superiores de Coacalco (TESCO). Se considera una base de datos de 232 personas como personal académico de las cuales se utilizan 180 para el cálculo de probabilidades, mismas que se prueban con el resto de la base de datos. Se calculan las probabilidades conjuntas y se utilizan para evaluar el clasificador. El modelo permite asignar probabilidades para cada uno de los nuevos prospectos de contratación de personal académico.

Palabras Clave - Bayes, Clasificador, Decisiones, Naive

Abstract - The Naive classifier model is used as a tool proposal to clarify the decision on the alternatives in the hiring of academic personnel of the Tecnológico de Estudios Superiores de Coacalco (TESCO). A database of 232 people is considered as academic personnel, of which 180 are used for the calculation of probabilities, which are tested with the rest of the database. The joint probabilities are calculated and used to evaluate the classifier. The model allows assigning probabilities for each of the new prospects for hiring academic staff.

Keywords - Bayes, Classifier, Decision, Naive.

Mathematical Subject Classification: 90B50.

I. INTRODUCCIÓN

El TESCO fue creado el 4 de septiembre de 1996, teniendo como objetivos [1]:

I. Formar profesionales e investigadores aptos para la aplicación y generación de conocimientos, con capacidad crítica y analítica en la solución de los problemas, con sentido innovador que incorpore los avances científicos y tecnológicos al ejercicio responsable de la profesión de acuerdo con los requerimientos del entorno, el estado y el país;

II. Realizar investigaciones científicas y tecnológicas que permitan el avance del conocimiento, el desarrollo de la enseñanza tecnológica y el mejor aprovechamiento social de los recursos naturales y materiales que contribuyan a la elevación de la calidad de vida comunitaria;

III. Colaborar con los sectores público, privado y social en la consolidación del desarrollo tecnológico y social de la comunidad;

IV. Realizar programas de vinculación con los sectores público, privado y social que contribuyan a la consolidación del desarrollo tecnológico y social del ser humano;

V. Realizar el proceso enseñanza-aprendizaje con actividades curriculares debidamente planeadas y ejecutadas; y

VI. Promover la cultura nacional y universal especialmente la de carácter tecnológico.

De los objetivos de creación del TESCO, se desprende la necesidad de contar con personal académico que coadyuve a lograr los objetivos planteados.

De acuerdo con [2] los tres métodos más usados en aprendizaje automático en tareas de clasificación son redes bayesianas, árboles de decisión y redes neuronales artificiales, además, como lo menciona [3] el objetivo principal de las técnicas de reconocimiento de formas, aplicadas a un problema general de clasificación, consiste en asignar a un objeto o fenómeno físico una de las diversas categorías o clases previamente especificadas.

También [3] indica que las redes bayesianas modelan un fenómeno mediante un conjunto de variables y las relaciones de dependencia entre ellas, con el modelo creado se pueden hacer inferencias bayesianas estimando las probabilidades posteriores de las variables no conocidas, con base a las variables conocidas.

Por su parte [4] menciona que el clasificador Naive -Bayes, es un modelo efectivo de clasificación por su simplicidad, resistencia al ruido, poco tiempo de procesamiento y alto poder predictivo. Además, se menciona que la tarea del *clasificador* es etiquetar (clasificar) correctamente un conjunto de datos en uno de los grupos o clases previamente definidas, así, un clasificador bayesiano es una función que asigna a un objeto u observación en la clase con mayor probabilidad.

Se mencionan algunas aplicaciones encontradas de los algoritmos de clasificación.

En [5] se aplica el método Naive Bayes a textos en redes sociales para clasificar los sentimientos sobre las elecciones; las categorías fueron (positivo, negativo o neutral). Se utilizan 50 datos de entrenamiento y 10 datos de prueba.

También en [6] se implementa minería de texto y sentimientos basados en documentos en datos de Twitter que son reprocesados a través de técnicas de aprendizaje automático utilizando el método Naive Bayes.

Por su parte [7], describen el uso del clasificador Naive Bayes para asignar probabilidades de pertenencia a individuos fuentes, según la ubicación de la fuente, las propiedades de rayos X y las propiedades visuales/infrarrojas.

Además, [8] utilizan el método Naive Bayes para la detección de intrusos en un sistema de Red Informático; el algoritmo implementado muestra una precisión del 84.42%.

Por último, en [9] se utiliza un enfoque propuesto llamado técnica Master-Slaves para mejorar la clasificación del texto árabe, se implementan cuatro clasificadores diferentes en el cuerpo recopilado; el clasificador Naive Bayes fue implementado como Maestro y los demás como Esclavos.

II. PROBLEMA

Por cuestiones de manejo de información, se desarrolla un cambio en el recuento de la base de datos del personal académico del TESCO. Se considera una plantilla de 232 trabajadores bajo la dirección académica. De acuerdo con [10], el personal docente se debe encontrar dentro de las categorías de:

- ✓ Profesor de Asignatura A
- ✓ Profesor de asignatura B
- ✓ Profesor de tiempo completo asociado A
- ✓ Profesor de tiempo completo asociado B
- ✓ Profesor de tiempo completo asociado C
- ✓ Profesor de tiempo completo Titular A

En la Tabla 1, se muestra la distribución del personal en cada una de las categorías.

Tabla I
Personal por categorías del personal académico.

Categoría	Porcentaje	Personal
Asignatura A	10%	23
Asignatura B	20%	46
Asociado A	30%	70
Asociado B	20%	46
Asociado C	15%	35
Titular A	5%	12

En la Tabla II se muestran las características curriculares del personal académico.

Tabla II
Características curriculares del personal académico.

Categoría	Características Curriculares
Experiencia profesional	Menos de 5 años
	Entre 5 y 10 años
	Mas de 10 años
Experiencia laboral	Menos de 5 años
	Entre 5 y 10 años
	Mas de 10 años
Título de licenciatura	Si
	No
Posgrado	Si
	No
Perfil deseable	Si
	No
SNI	Si
	No

Debido al cambio constante de personal académico, se requiere una herramienta que permita dar claridad en la

decisión sobre la contratación del nuevo personal académico que ayude en la asignación de la categoría correspondiente de acuerdo con las características curriculares del nuevo personal.

III. DESARROLLO

En la Figura 1 se muestra la estructura para el clasificador del problema planteado.

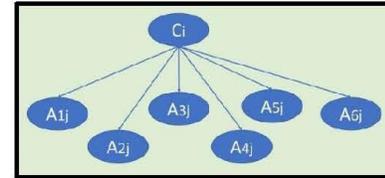


Fig. 1. Estructura para clasificador

Donde

C_i = Categorías
 $i = 1$, Asignatura A
 $i = 2$, Asignatura B
 $i = 3$, Asociado A
 $i = 4$, Asociado B
 $i = 5$, Asociado C
 $i = 6$, Titular A

a_{1j} = Experiencia Docente
 $j = 1$, Menos de 5 años
 $j = 2$, Entre 5 y 10 años
 $j = 3$, Mas de 10 años

a_{2j} = Experiencia Profesional
 $j = 1$, Menos de 5 años
 $j = 2$, Entre 5 y 10 años
 $j = 3$, Mas de 10 años

a_{3j} = Título de licenciatura
 $j = 1$, Si
 $j = 2$, No

a_{4j} = Posgrado
 $j = 1$, Si
 $j = 2$, No

a_{5j} = Perfil deseable
 $j = 1$, Si
 $j = 2$, No

a_{6j} = SNI
 $j = 1$, Si
 $j = 2$, No

De acuerdo con el Teorema de Bayes, la probabilidad de que una observación $(a_{1j}, a_{2j}, a_{3j}, a_{4j}, a_{5j}, a_{6j})$ pertenezca a la clase C_i se calcula con la ecuación 1.

$$P(C = c_i | a_{1i}, a_{2i}, a_{3i}, a_{4i}, a_{5i}, a_{6i}) = \frac{P(C=c)P(a_{1i}, a_{2i}, a_{3i}, a_{4i}, a_{5i}, a_{6i} | C=c)}{P(a_{1i}, a_{2i}, a_{3i}, a_{4i}, a_{5i}, a_{6i})} \quad (1)$$

Se consideran 180 elementos de la plantilla adscrita a la dirección académica para desarrollar las distribuciones de probabilidad, se estiman los parámetros de las distribuciones de probabilidad que se describen en la red de la Figura 1.

En la Tabla III se muestran los resultados de los datos curriculares observados.

Tabla III
Datos curriculares observados

Experiencia Docente	ASIGNATURA A	ASIGNATURA B	ASOCIADO A	ASOCIADO B	ASOCIADO C	TITULAR A
menos de 5 años	7	19	23	5	29	4
entre 5 y 10 años	5	7	19	6	21	3
mas de 10 años	1	9	10	4	4	4
TOTALES	13	35	52	15	54	11
Experiencia Profesional	ASIGNATURA A	ASIGNATURA B	ASOCIADO A	ASOCIADO B	ASOCIADO C	TITULAR A
menos de 5 años	5	20	31	8	28	6
entre 5 y 10 años	5	11	19	4	17	3
mas de 10 años	3	4	2	3	9	2
TOTALES	13	35	52	15	54	11
TITULO LICENCIATURA	ASIGNATURA A	ASIGNATURA B	ASOCIADO A	ASOCIADO B	ASOCIADO C	TITULAR A
SI	13	35	52	15	54	11
NO	0	0	0	0	0	0
TOTALES	13	35	52	15	54	11
POSGRADO	ASIGNATURA A	ASIGNATURA B	ASOCIADO A	ASOCIADO B	ASOCIADO C	TITULAR A
SI	0	0	40	11	45	8
NO	13	35	12	4	9	3
TOTALES	13	35	52	15	54	11
PERFIL DESEABLE	ASIGNATURA A	ASIGNATURA B	ASOCIADO A	ASOCIADO B	ASOCIADO C	TITULAR A
SI	0	0	44	9	43	7
NO	13	35	8	6	11	4
TOTALES	13	35	52	15	54	11
SNII	ASIGNATURA A	ASIGNATURA B	ASOCIADO A	ASOCIADO B	ASOCIADO C	TITULAR A
SI	0	0	0	0	0	4
NO	13	35	52	15	54	7
TOTALES	13	35	52	15	54	11

Considerando que la estimación de la distribución de probabilidad de la categoría $P(C_i)$ y de la distribución condicional $P(a_{ij} | C_i)$ se basa en la frecuencia relativa de ocurrencia de los correspondientes sucesos, el valor para las probabilidades estimadas será:

$$P(a_{ij} | C_i) = \frac{n(a_{ij}, C=c)}{n(C=c)} \quad (2)$$

De la Tabla III se puede observar que todos los casos de la característica curricular a_{32} (Titulo de licenciatura = no) no cuenta con ningún elemento motivo por el cual se decide eliminar dicha categoría. También se observa que algunas características curriculares no cuentan con elementos, pero no en todos los casos, por lo que se aplica el corrector de Laplace [11], el cual evita que se asigne una probabilidad nula para que no se elimine por completo la probabilidad final.

Aplicando el corrector de Laplace y la ecuación 2 a la Tabla III, se tienen los resultados de las probabilidades conjuntas, que se muestran en la Tabla IV.

Tabla IV.
Probabilidades Conjuntas

P(A1 Ci)	P(Requisito Experiencia Docente)		
	menos de 5 años	entre 5 y 10 años	más de 10 años
Asignatura A	1/2	3/8	1/8
Asignatura B	10/19	4/19	5/19
Asociado A	24/55	4/11	1/5
Asociado B	1/3	7/18	5/18
Asociado C	10/19	22/57	5/57
Titular A	4/11	3/11	4/11

P(A2 Ci)	P(Requisito Experiencia Profesional)		
	menos de 5 años	entre 5 y 10 años	más de 10 años
Asignatura A	3/8	3/8	1/4
Asignatura B	21/38	6/19	5/38
Asociado A	32/55	4/11	3/55
Asociado B	1/2	5/18	2/9
Asociado C	29/57	6/19	10/57
Titular A	6/11	3/11	2/11

P(A3 Ci)	P(Posgrado)	
	SI	No
Asignatura A	1/8	7/8
Asignatura B	1/19	18/19
Asociado A	10/13	3/13
Asociado B	11/15	4/15
Asociado C	5/6	1/6
Titular A	8/11	3/11

P(A4 Ci)	P(Perfil Deseable)	
	SI	No
Asignatura A	1/8	7/8
Asignatura B	1/37	36/37
Asociado A	11/13	2/13
Asociado B	3/5	2/5
Asociado C	43/54	11/54
Titular A	7/11	4/11

P(A5 Ci)	P(SNII)	
	SI	No
Asignatura A	1/8	7/8
Asignatura B	1/19	18/19
Asociado A	2/55	53/55
Asociado B	1/9	8/9
Asociado C	2/57	55/57
Titular A	4/11	7/11

De igual forma, de la misma Tabla III se obtienen la distribución de probabilidad para cada una de las categorías que se muestran en la Tabla V.

Tabla V
Probabilidades por Categoría

P(CATEGORIA j)	ASIGNATURA A	ASIGNATURA B	ASOCIADO A	ASOCIADO B	ASOCIADO C	TITULAR A
	A	B	A	B	C	A
	16	38	55	18	57	11
	16/195	38/195	11/39	6/65	19/65	11/195

En la Tabla VI se muestran las 72 combinaciones posibles para las características curriculares del personal académico del TESCO.

IV. RESULTADOS

Con los resultados de las probabilidades conjuntas que se muestran en la Tabla IV se prueba el clasificador, evaluando las probabilidades para los 52 elementos de la base de datos que no se ocuparon para el desarrollo de este.

Tabla VI
Características Curriculares

Se ilustra el cálculo para la alternativa

$$\alpha P(C_i)P(a_{11}|C_i)P(a_{21}|C_i)P(a_{31}|C_i)P(a_{41}|C_i)P(a_{51}|C_i)P(a_{61}|C_i) = \alpha P(C_i)P(a_{11}|C_i)P(a_{21}|C_i)P(a_{31}|C_i)P(a_{41}|C_i)P(a_{51}|C_i)P(a_{61}|C_i) \quad (3)$$

Donde α es la constancia de proporcionalidad.

Sustituyendo la ecuación (3) para cada una de las alternativas se obtiene

$$\alpha \left(\frac{16}{195}\right) \left(\frac{1}{2}\right) \left(\frac{3}{8}\right) \left(\frac{1}{8}\right) \left(\frac{1}{8}\right) \left(\frac{1}{8}\right) = \frac{1}{33280} \alpha \quad (4)$$

$$\alpha \left(\frac{38}{195}\right) \left(\frac{10}{19}\right) \left(\frac{21}{38}\right) \left(\frac{1}{19}\right) \left(\frac{1}{37}\right) \left(\frac{1}{19}\right) = \frac{14}{3299179} \alpha \quad (5)$$

$$\alpha \left(\frac{11}{39}\right) \left(\frac{24}{55}\right) \left(\frac{32}{55}\right) \left(\frac{10}{13}\right) \left(\frac{11}{13}\right) \left(\frac{2}{55}\right) = 0.0017 \alpha \quad (6)$$

$$\alpha \left(\frac{6}{65}\right) \left(\frac{1}{3}\right) \left(\frac{1}{2}\right) \left(\frac{11}{15}\right) \left(\frac{3}{5}\right) \left(\frac{1}{9}\right) = \frac{11}{14625} \alpha \quad (7)$$

$$\alpha \left(\frac{19}{65}\right) \left(\frac{10}{19}\right) \left(\frac{29}{57}\right) \left(\frac{5}{6}\right) \left(\frac{43}{54}\right) \left(\frac{2}{57}\right) = 0.0018 \alpha \quad (8)$$

$$\alpha \left(\frac{11}{195}\right) \left(\frac{4}{11}\right) \left(\frac{6}{11}\right) \left(\frac{8}{11}\right) \left(\frac{7}{11}\right) \left(\frac{4}{11}\right) = 0.0019 \alpha \quad (9)$$

Normalizando las ecuaciones anteriores.

$$\alpha \left(\frac{1}{33280} + \frac{14}{3299179} + 0.0017 + \frac{11}{14625} + 0.0018 + 0.0019\right) = 1 \quad (10)$$

Por lo tanto

$$\alpha = 161.644 \quad (11)$$

Donde α es el valor necesario para que se cumpla la ecuación (10).

	Experiencia Docente	Experiencia Profesional	Posgrado	Perfil Deseable	SNII
1	x < 5	x < 5	SI	SI	SI
2	5 < x < 10	x < 5	SI	SI	SI
3	x > 10	x < 5	SI	SI	SI
4	x < 5	5 < x < 10	SI	SI	SI
5	5 < x < 10	5 < x < 10	SI	SI	SI
6	x > 10	5 < x < 10	SI	SI	SI
7	x < 5	x > 10	SI	SI	SI
8	5 < x < 10	x > 10	SI	SI	SI
9	x > 10	x > 10	SI	SI	SI
10	x < 5	x < 5	NO	SI	NO
11	5 < x < 10	x < 5	NO	SI	NO
12	x > 10	x < 5	NO	SI	NO
13	x < 5	5 < x < 10	NO	SI	NO
14	5 < x < 10	5 < x < 10	NO	SI	NO
15	x > 10	5 < x < 10	NO	SI	NO
16	x < 5	x > 10	NO	SI	NO
17	5 < x < 10	x > 10	NO	SI	NO
18	x > 10	x > 10	NO	SI	NO
19	x < 5	x < 5	SI	NO	SI
20	5 < x < 10	x < 5	SI	NO	SI
21	x > 10	x < 5	SI	NO	SI
22	x < 5	5 < x < 10	SI	NO	SI
23	5 < x < 10	5 < x < 10	SI	NO	SI
24	x > 10	5 < x < 10	SI	NO	SI
25	x < 5	x > 10	SI	NO	SI
26	5 < x < 10	x > 10	SI	NO	SI
27	x > 10	x > 10	SI	NO	SI
28	x < 5	x < 5	NO	NO	NO
29	5 < x < 10	x < 5	NO	NO	NO
30	x > 10	x < 5	NO	NO	NO
31	x < 5	5 < x < 10	NO	NO	NO
32	5 < x < 10	5 < x < 10	NO	NO	NO
33	x > 10	5 < x < 10	NO	NO	NO
34	x < 5	x > 10	NO	NO	NO
35	5 < x < 10	x > 10	NO	NO	NO
36	x > 10	x > 10	NO	NO	NO
37	x < 5	x < 5	SI	SI	SI
38	5 < x < 10	x < 5	SI	SI	SI
39	x > 10	x < 5	SI	SI	SI
40	x < 5	5 < x < 10	SI	SI	SI
41	5 < x < 10	5 < x < 10	SI	SI	SI
42	x > 10	5 < x < 10	SI	SI	SI
43	x < 5	x > 10	SI	SI	SI
44	5 < x < 10	x > 10	SI	SI	SI
45	x > 10	x > 10	SI	SI	SI
46	x < 5	x < 5	NO	SI	NO
47	5 < x < 10	x < 5	NO	SI	NO
48	x > 10	x < 5	NO	SI	NO
49	x < 5	5 < x < 10	NO	SI	NO
50	5 < x < 10	5 < x < 10	NO	SI	NO
51	x > 10	5 < x < 10	NO	SI	NO
52	x < 5	x > 10	NO	SI	NO
53	5 < x < 10	x > 10	NO	SI	NO
54	x > 10	x > 10	NO	SI	NO
55	x < 5	x < 5	SI	NO	SI
56	5 < x < 10	x < 5	SI	NO	SI
57	x > 10	x < 5	SI	NO	SI
58	x < 5	5 < x < 10	SI	NO	SI
59	5 < x < 10	5 < x < 10	SI	NO	SI
60	x > 10	5 < x < 10	SI	NO	SI
61	x < 5	x > 10	SI	NO	SI
62	5 < x < 10	x > 10	SI	NO	SI
63	x > 10	x > 10	SI	NO	SI
64	x < 5	x < 5	NO	NO	NO
65	5 < x < 10	x < 5	NO	NO	NO
66	x > 10	x < 5	NO	NO	NO
67	x < 5	5 < x < 10	NO	NO	NO
68	5 < x < 10	5 < x < 10	NO	NO	NO
69	x > 10	5 < x < 10	NO	NO	NO
70	x < 5	x > 10	NO	NO	NO
71	5 < x < 10	x > 10	NO	NO	NO
72	x > 10	x > 10	NO	NO	NO

Con el valor de α mostrado en la ecuación 11 se obtienen las siguientes probabilidades para cada una de las categorías dependiendo de las características curriculares de la ecuación 3

$$P(\text{Asignatura A} | a_{i1}) = 0.006 \quad (12)$$

$$P(\text{Asignatura B} | a_{i1}) = 0.002 \quad (13)$$

$$P(\text{Asociado A} | a_{i1}) = 0.274 \quad (14)$$

$$P(\text{Asociado B} | a_{i1}) = 0.122 \quad (15)$$

$$P(\text{Asociado C} | a_{i1}) = 0.290 \quad (15)$$

$$P(\text{Titular A} | a_{i1}) = 0.306 \quad (16)$$

Con el procedimiento descrito en las ecuación. 3 a la 16 se aplica a los 52 datos de prueba, obteniendo los resultados que se muestran en la Tabla VII.

Tabla VII
Evaluación del Clasificador

Categoría	Elementos de Prueba	Clasificados Correctamente	Porcentaje
Asignatura A	4	2	50%
Asignatura B	10	4	40%
Asociado A	15	9	60%
Asociado B	4	3	75%
Asociado C	15	8	53%
Titular A	4	2	50%

V. CONCLUSIONES

Se puede observar que el clasificador cuenta con una eficacia promedio superior al 50%, sobre todo en las categorías de Asociado A, B y C. Para que mejore la eficacia del clasificador se debe pedir que se analicen otras características curriculares que permitan dar mayor claridad en la evaluación de prospectos de personal académico para el TESCO.

REFERENCIAS

- [1] Gobierno del Estado de México, "Gaceta de Gobierno", México, 1996. Consultado: oct. 03, 2022. [En línea]. Disponible en: https://legislacion.edomex.gob.mx/ve_periodico_oficial
- [2] C. Malagón Luque, "Clasificadores bayesianos. El algoritmo Naive Bayes". Consultado: oct. 03, 2022. [En línea]. Disponible http://www.nebrija.es/~cmalagon/inco/Apuntes/bayesian_learning.pdf
- [3] B. Sierra Araujo, "Aprendizaje automático: conceptos básicos y avanzados", Primera., vol. 1. España: Pearson - Prentice Hall, 2006.
- [4] F. de, C. Matemáticas, U. de Posgrado, A. Carlos, y A. Jaimes, "Optimización del clasificador 'Naive Bayes' usando árbol de decisión c4.5" tesis para optar el grado académico de magister en estadística, Universidad Nacional Mayor de San Marcos. 2015.
- [5] D. Nurcahyono, W. P. Putra, A. Najib, y T. R. Tulili, "Analysis sentiment in social media against election using the method naive Bayes", en Journal of Physics: Conference Series, jun. 2020, vol. 1511, núm. 1. doi: 10.1088/1742-6596/1511/1/012003.
- [6] Samsir et al., "Naives Bayes Algorithm for Twitter Sentiment Analysis", en Journal of Physics: Conference Series, jun. 2021, vol. 1933, núm. 1. doi: 10.1088/1742-6596/1933/1/012019.
- [7] P. S. Broos, K. v. Getman, M. S. Povich, L. K. Townsley, E. D. Feigelson, y G. P. Garmire, "A naive bayes source classifier for X-ray sources", Astrophysical Journal, Supplement Series, vol. 194, núm. 1, may 2011, doi: 10.1088/0067-0049/194/1/4.
- [8] P. A. Rahayuningsih, R. Maulana, W. Imayani, D. Saputra, y D. Purwaningtiyas, "Feature Dependent Naive Bayes for Network Intrusion

Detection System", en Journal of Physics: Conference Series, nov. 2020, vol. 1641, núm. 1. doi: 10.1088/1742-6596/1641/1/012023.

- [9] Z. A. Abutiheen, A. H. Aliwy, y K. B. S. Aljanabi, "Arabic text classification using master-slaves technique", en Journal of Physics: Conference Series, jun. 2018, vol. 1032, núm. 1. doi: 10.1088/1742-6596/1032/1/012052.
- [10] Tecnológico nacional de México, "Lineamientos que establecen los requisitos y condiciones que debe reunir el personal académico que imparte planes y programas de estudio en los institutos tecnológicos descentralizados", Consultado: oct. 03, 2022. [En línea]. Disponible en: <http://www.dgest.gob.mx/itd/academica-ditd>.
- [11] TeamUNO blog A project from a data scientists team, "Naive Bayes como herramienta contra la agregación y segregación de datos", Consultado: oct. 03, 2022. [En línea]. Disponible en: <http://teamuno.github.io/2015/07/12/uso-naive-bayes/#:~:text=Recordemos%20que%20la%20correcci%C3%B3n%20de,por%20completo%20la%20probabilidad%20final,> nov. 02, 2022.

13.2 Producto Académico 2



ÁRBOL CLASIFICADOR SUPERVISADO COMO APOYO A LA DECISIÓN EN LA CONTRATACIÓN DE PERSONAL ACADÉMICO DEL TESCO.

Viveros-Rosas Leopoldo, TECNM – TESCO.
leopoldo@tesco.edu.mx
Díaz-Téllez Rebeca, TECNM – TESCO.
rebeca.sub.a@tesco.edu.mx
Chew-Hernández Mario Luis, TECNM – TESCO.
mario@tesco.edu.mx
Vega-González José Antonio, TECNM – TESCO.
a.vega@tesco.edu.mx

Recibido: marzo de 2023

Aceptado: mayo de 2023

RESUMEN

Considerando que los árboles de decisión son de los algoritmos más sencillos, pero también los más poderosos del *Machine Learning*, en este trabajo se presenta una aplicación de las herramientas para la generación de árboles de decisión, ya que se cuenta con *set* de datos complejos. En especial se desarrolla un árbol clasificador para el personal bajo la dirección académica del Tecnológico de Estudios Superiores de Coacalco (TESCo). El objetivo es identificar características relevantes para la contratación del personal académico y desarrollar el árbol de decisión correspondiente. Se considera una plantilla de 232 elementos, utilizando 180 de estos para desarrollar el árbol, los 52 restantes se utilizan para medir la eficacia del árbol. Aun cuando se ha modificado la información por cuestiones de privacidad el algoritmo queda desarrollado para alimentarlo con los datos sin modificación. Se muestra el árbol generado cuya medida de calidad alcanza el 60% de clasificaciones correctas.

ABSTRACT

Considering that decision trees are one of the simplest algorithms, but also the most powerful in Machine Learning, this paper presents an application of the tools for the generation of decision trees, since it has a complex data set. In particular, a classifying tree is developed for the personnel under the academic direction of the Tecnológico de Estudios Superiores de Coacalco (TESCo). The objective is to identify relevant characteristics for the recruitment of academic staff and develop the corresponding decision tree. A template of 232 elements is considered, using 180 of these to develop the tree, the remaining 52 are used to measure the effectiveness of the tree. Even when the information has been modified for privacy reasons, the algorithm is developed to feed it with the data without modification. The generated tree whose quality measure reaches 60% of correct classifications is shown.

PALABRAS CLAVE: Algoritmo clasificador, Árboles de decisión, Apoyo a la decisión.

1 INTRODUCCIÓN

El TESCO fue creado el 4 de septiembre de 1996, teniendo como objetivos [1]

- I. Formar profesionales e investigadores aptos para la aplicación y generación de conocimientos, con capacidad crítica y analítica en la solución de los problemas, con sentido innovador que incorpore los avances científicos y tecnológicos al ejercicio responsable de la profesión de acuerdo con los requerimientos del entorno, el estado y el país;
- II. Realizar investigaciones científicas y tecnológicas que permitan el avance del conocimiento, el desarrollo de la enseñanza tecnológica y el mejor aprovechamiento social de los recursos naturales y materiales que contribuyan a la elevación de la calidad de vida comunitaria;
- III. Colaborar con los sectores público, privado y social en la consolidación del desarrollo tecnológico y social de la comunidad;
- IV. Realizar programas de vinculación con los sectores público, privado y social que contribuyan a la consolidación del desarrollo tecnológico y social del ser humano;
- V. Realizar el proceso enseñanza-aprendizaje con actividades curriculares debidamente planeadas y ejecutadas; y
- VI. Promover la cultura nacional y universal especialmente la de carácter tecnológico.

De los objetivos de creación del TESCO, se desprende la necesidad de contar con personal académico que coadyuve a lograr los objetivos planteados.

De acuerdo con [2] los tres métodos más usados en aprendizaje automático en tareas de clasificación son redes bayesianas, árboles de decisión y redes neuronales



artificiales, además, como lo menciona [3] el objetivo principal de las técnicas de reconocimiento de formas, aplicadas a un problema general de clasificación, consiste en asignar a un objeto o fenómeno físico una de las diversas categorías o clases previamente especificadas.

La clasificación supervisada enfoca el problema clasificatorio partiendo de un conjunto de objetos descritos por un vector de características y a la clase a la que pertenecen cada uno de ellos, a este conjunto se le conoce como conjunto de entrenamiento o conjunto de aprendizaje [3].

También en [3] se menciona que la clasificación supervisada ha sido utilizada en numerosos problemas de distinta índole como el diagnóstico de enfermedades, la concesión o rechazo de créditos de banca, predicción de quiebra o bancarrota en empresas, reconocimiento de caracteres escritos a mano, detección de anomalías en cromosomas, desambigüedades de calles mal escritas, entre otras; se mencionan algunas de aplicaciones revisadas.

En [4] se estudia la eficiencia de la clasificación de estrellas/galaxias de diferentes algoritmos de árboles de decisión. Cada algoritmo está definido por un conjunto de parámetros que, cuando varían, producen diferentes árboles de clasificación final.

Por su parte, en [5] se presenta un algoritmo basado en la técnica de aprendizaje automático de árboles de decisión potenciados. El método presentado conduce a la reconstrucción de una función de densidad de probabilidad completa para el corrimiento al rojo fotométrico de cada galaxia y también proporciona una figura de mérito de calidad foto-z para cada galaxia, galaxia que se puede utilizar para rechazar valores atípicos.

En [6] se menciona que los árboles de decisión son un modelo de clasificación utilizado en la inteligencia artificial, cuya principal característica es su aporte visual a la toma de decisiones. Evalúan el desempeño de tres algoritmos para la construcción de estos árboles utilizando dos bases de datos que contienen informes médicos de pacientes reales. Estos datos corresponden a la sintomatología que un médico especialista considera para el diagnóstico de cáncer de seno.

Además, en [7] se menciona que los algoritmos de árbol de decisiones son algoritmos de aprendizaje inductivo basado en información preestablecida que infieren una regla de clasificación representada por una estructura de árbol a través del aprendizaje del conjunto de entrenamiento. Comentan que el uso de algoritmos de árboles de decisión para realizar la extracción de datos sobre la información del usuario de telefonía móvil puede analizar con precisión las necesidades del usuario y la

satisfacción del usuario con los servicios. Su investigación realiza un análisis y una extracción de datos de usuarios móviles para mejorar el nivel de los servicios de marketing móvil e implementar estrategias de gestión refinadas de clientes.

Por otro lado, en [8] se emplea el algoritmo C4.5 para la creación de árboles de decisión para ayudar a los clubes pequeños de la Serie A italiana a encontrar los objetivos mínimos para evitar el descenso a la competición de la Serie B. Los métodos utilizados en el estudio son la recopilación de datos, el preprocesamiento de datos, la propuesta de modelos, las pruebas de modelos y la validación de modelos, esperando que el valor de precisión supere el 85%.

En [9] se menciona que la clasificación automatizada de los contaminantes del aire ambiental es una tarea importante en la evaluación de los peligros de la contaminación del aire y la investigación de la calidad de vida. En su estudio describen varios tipos de algoritmos de árbol de decisiones para encontrar la intercorrelación entre el índice de contaminación del aire (API) dominante para los valores del percentil de PM10 y otros cuatro contaminantes del aire, además de otros dos parámetros meteorológicos: temperatura ambiente y humedad.

En [10] se comenta que los árboles de decisión son una implementación de sistemas desarrollados por las personas en la búsqueda y toma de decisiones para problemas de decisión que toman en cuenta diversos factores relacionados dentro del alcance del problema. Una de las conclusiones de su investigación es que el método de árbol de decisión tiene una alta velocidad en la clasificación de las predicciones de rendimiento de los estudiantes. Durante el procesamiento, el método Árbol de decisión produce una clasificación de árboles de decisión basada en el valor promedio del final del semestre para que pueda conocer el porcentaje/comparación de aumentos o disminuciones del rendimiento de los estudiantes.

En [11] se utiliza el algoritmo del árbol de decisiones para realizar una investigación sobre la evaluación de riesgos de los clientes antes de préstamo de crédito monetario. Después de la verificación del modelo del método de validación cruzada y el método de validación aleatoria, los resultados muestran que la precisión de la estimación del riesgo del algoritmo del árbol de decisión alcanza el 81.2% y el 83.6 %, demostrando que los modelos de árbol de decisión pueden ser una referencia eficaz para la evaluación de riesgos.

En [12] se menciona que los árboles de decisión con enfoques de algoritmos C5.0 y J48 se utilizan para clasificar enormes cantidades de datos y en su estudio los utilizan para encontrar la causa raíz de los retrasos en los almacenes, encontrando que el cálculo de la ganancia de



información muestra que las variables que más influyen para el proceso de clasificación de la demora son la capacidad del almacén y el tiempo de entrega de la orden de cotización.

En [13] Desarrollan un árbol de decisión utilizando el algoritmo C4.5 para determinar la clasificación de los becarios. El algoritmo tiene la ventaja de la velocidad de información y lectura de modelos, alta precisión, flexibilidad y procesamiento de datos discretos y continuos. En el estudio utilizan 460 datos y para cada atributo se la ganancia de información permitiendo producir árboles y reglas de decisión. Los escenarios se utilizan como referencia para la clasificación de nuevos candidatos ya que se cuenta con una precisión más alta de 85.55%.

En [14] se compara el rendimiento de los métodos de árboles de decisión en el reconocimiento de la actividad humana utilizando datos de aceleración y sacudidas. Los sujetos realizaron actividades humanas de la vida diaria, a saber, caminar sobre una superficie plana, subir y bajar escaleras, sentarse, pararse y acostarse. Sus resultados mostraron que Random Forest superó a los otros clasificadores con funciones de aceleración que funcionaron mejor que las funciones de sacudida. Sin embargo, las características combinadas de aceleración y tirón producen la mayor precisión. En conclusión, Random Forest es la mejor técnica de árboles de decisión para reconocer el patrón en la actividad humana.

2 PROBLEMA

Por cuestiones de manejo de información, se desarrolla un cambio en el recuento de la base de datos del personal académico del TESCO. Se considera una plantilla de 232 trabajadores bajo la dirección académica. De acuerdo con [15], el personal docente se debe encontrar dentro de las categorías de:

- Profesor de Asignatura A
- Profesor de asignatura B
- Profesor de tiempo completo asociado A
- Profesor de tiempo completo asociado B
- Profesor de tiempo completo asociado C
- Profesor de tiempo completo Titular A

En la Tabla 1, se muestra la distribución del personal en cada una de las categorías.

Tabla 1. Personal por categorías del personal académico

Categoría	Porcentaje	Personal
Asignatura A	10%	23
Asignatura B	20%	46

Asociado A	30%	70
Asociado B	20%	46
Asociado C	15%	35
Titular A	5%	12

En la Tabla 2 se muestran las características curriculares del personal académico del TESCO.

Tabla 2. Características curriculares del personal académico del TESCO.

Categoría	Características Curriculares
Experiencia profesional	Menos de 5 años
	Entre 5 y 10 años
	Mas de 10 años
Experiencia laboral	Menos de 5 años
	Entre 5 y 10 años
	Mas de 10 años
Título de licenciatura	Si
	No
Posgrado	Si
	No
Perfil deseable	Si
	No
SNI	Si
	No

Debido al cambio constante de personal académico, se requiere una herramienta que permita dar claridad en la decisión sobre la contratación del nuevo personal académico que ayude en la asignación de la categoría correspondiente de acuerdo con las características curriculares del nuevo personal.

3 MÉTODO

De acuerdo con [16] El aprendizaje automático es una disciplina de la Inteligencia Artificial cuyo objetivo principal es que un agente artificial mejore su rendimiento (aprenda) a partir de resultados previamente obtenidos. Desde el punto de vista del aprendizaje automático, obtener conocimiento implica extraer información potencialmente útil y desconocida a partir de un conjunto de datos.

Para el aprendizaje supervisado existen dos técnicas principales: clasificación y regresión. Las tareas de clasificación requieren un vector de n-datos de entrada y la asignación de una de las k-clases discretas. [16]

En la Figura 1 se ilustra el proceso general de clasificación.

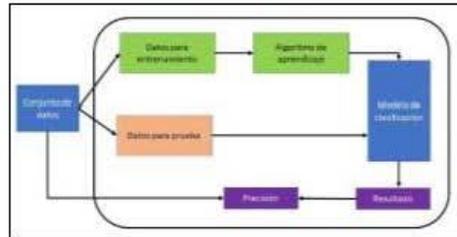


Figura 1. Proceso de clasificación.

También en [16] se dice que el árbol de decisión es una estructura jerárquica que clasifican un conjunto de datos, al clasificarlos desde el nodo o vértice raíz hasta algún vértice final. Cada vértice interno en el árbol especifica una condición de prueba que evalúa uno o varios atributos, y cada rama o arista descendiente del árbol representa una secuencia de decisiones hechas por el modelo para determinar la membresía o pertenencia de la clase de un nuevo ejemplo no clasificado. El aprendizaje o la inducción de árboles de decisión se realiza a partir de un conjunto de datos mediante un algoritmo que permite obtener un modelo que intenta descubrir la relación que existe entre los atributos de entrada y la clase objetivo. Al aplicarlo a nuevos ejemplos este es capaz de predecir el valor de la clase objetivo y de esta manera clasificarlos correctamente.

Para formar el árbol de decisión se calculan los valores de entropía y ganancia de la información descritos en [17].

$$Entropía P = -\sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

Entendiendo como entropía como incertidumbre existente en la cuenta de información.

$$Ganancia (p,T) = Entropía P - \sum_{j=1}^n p_j Entropía(p_j) \quad (2)$$

Entendiendo la ganancia como un decremento de la entropía/incertidumbre del sistema.

De esta forma se desarrolla el árbol de decisión que permita un a menor incertidumbre en el sistema.

4 DESARROLLO

Se consideran 180 elementos de la plantilla adscrita a la dirección académica para desarrollar las distribuciones de probabilidad que se muestran en la Tabla 3.

Tabla 3. Datos curriculares observados.

Experiencia Docente	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Título A
menos de 5 años	7	19	23	5	29	4
entre 5 y 10 años	5	7	19	6	21	3
mas de 10 años	1	9	10	4	4	4
Experiencia Profesional	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Título A
menos de 5 años	5	20	31	8	28	6
entre 5 y 10 años	5	11	19	4	17	3
mas de 10 años	3	4	2	3	9	2
Posgrado	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Título A
SI	0	0	40	11	45	8
No	13	35	12	4	9	3
Perfil Deseable	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Título A
SI	0	0	44	3	43	7
No	13	35	9	6	11	4
SNM	Asignatura A	Asignatura B	Asociado A	Asociado B	Asociado C	Título A
SI	0	0	0	0	0	4
No	13	35	32	15	54	7
Totales	13	35	32	15	54	11

De acuerdo con la ecuación (1) se hace el cálculo de la entropía para la clase *tipo de plaza*.

$$P(Plaza) = 2.3170$$

Cálculo de la entropía para cada uno de los posibles valores de la variable *Experiencia Docente*.

$$P(plaza|ED < 5 años) = 2.2487$$

$$P(plaza|5 años < ED < 10 años) = 2.2508$$

$$P(plaza|ED > 10 años) = 2.3204$$

De acuerdo con (2) se calcula la Ganancia de información para la variable *Experiencia Docente*.

$$G(ED) = 0.0548$$

Repetiendo el proceso para cada una de las variables, se tiene que

$$G(EP) = 0.0339$$

$$G(P) = 0.4409$$

$$G(PD) = 0.4485$$

$$G(SNM) = 0.0960$$

Se puede observar que la mejor ganancia de información se obtiene con la variable de *Perfil Deseable*. Por lo tanto, el nodo o vértice raíz del árbol de decisión será *Perfil Deseable*.

Ahora, se calcula la entropía para la clase *Plaza* condicionado a cada uno de los resultados de la variable *Perfil Deseable*.

$$P(Plaza|Perfil Deseable = si) = 1.6212$$



$$P(\text{Plaza}|\text{Perfil Deseable} = \text{no}) = 2.1993$$

Cuyos cálculos de ganancia de información son:

$$G(ED|PD = \text{si}) = 0.0353$$

$$G(EP|PD = \text{si}) = 0.0700$$

$$G(P|PD = \text{si}) = 0.0492$$

$$G(SNI|PD = \text{si}) = 0.1700$$

$$G(ED|PD = \text{no}) = 0.6310$$

$$G(EP|PD = \text{no}) = 1.1431$$

$$G(P|PD = \text{no}) = 1.7712$$

$$G(SNI|PD = \text{no}) = 2.1993$$

Con los valores obtenidos de Ganancia de información se observa que en ambos casos cuando la variable *Perfil Deseable* toma el valor de *si* o *no*, el siguiente nodo o vértice debe ser la variable *SNI*.

Se desarrollan los cálculos de Ganancia de Información para cada una de las opciones de las variables *Perfil Deseable* y *SNI*

$$P(\text{Plaza}|PD = \text{si} \cap SNI = \text{si}) = 0$$

$$P(\text{Plaza}|PD = \text{si} \cap SNI = \text{no}) = 1.5099$$

$$G(ED|PD = \text{si}, SNI = \text{no}) = 0.0750$$

$$G(EP|PD = \text{si}, SNI = \text{no}) = 0.0738$$

$$G(P|PD = \text{si}, SNI = \text{no}) = 0.0488$$

$$P(\text{Plaza}|PD = \text{no} \cap SNI = \text{si}) = 0$$

$$P(\text{Plaza}|PD = \text{no} \cap SNI = \text{no}) = 2.1993$$

$$G(ED|PD = \text{no}, SNI = \text{no}) = 0.6310$$

$$G(EP|PD = \text{no}, SNI = \text{no}) = 1.1431$$

$$G(P|PD = \text{no}, SNI = \text{no}) = 1.7712$$

Se trabaja la arista o rama de la variable *Posgrado* cuando $PD = \text{no}$, $SNI = \text{no}$.

$$P(\text{Plaza}|PD = \text{no} \cap SNI = \text{no} \cap P = \text{si}) = 1.5219$$

$$P(\text{Plaza}|PD = \text{no} \cap SNI = \text{no} \cap P = \text{no}) = 2.1017$$

$$G(ED|PD = \text{no}, SNI = \text{no}, P = \text{si}) = 0.7219$$

$$G(EP|PD = \text{no}, SNI = \text{no}, P = \text{si}) = 0$$

$$G(ED|PD = \text{no}, SNI = \text{no}, P = \text{no}) = 0.7090$$

$$G(EP|PD = \text{no}, SNI = \text{no}, P = \text{no}) = 1.1995$$

Se puede observar que para el caso de que la variable *Posgrado* tome un valor de *si*, el nodo o vértice siguiente será *Experiencia Docente*, de lo contrario se usa la variable *Experiencia profesional*.

Con los cálculos de ganancia se desarrolla el árbol que se muestra en la sección de resultados.

5 RESULTADOS

En la Figura 1 se muestra la parte del árbol de decisión que corresponde a la primera alternativa del nodo o vértice raíz cuando *Perfil Deseable* = *si*.

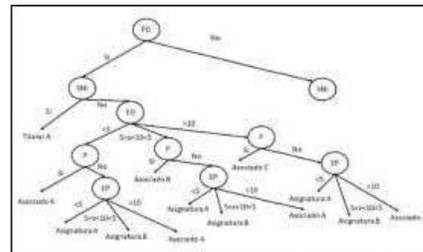


Figura 2. Sección del árbol de decisión cuando $PD = \text{si}$.

De la Figura 2 se puede observar que para que se obtenga una clasificación con nivel de *Asociado* se requiere que la variable *Posgrado* tome el valor de *si*, o que *Experiencia Profesional* y *Experiencia Docente* tomen sus valores más altos.

En la Figura 3 se muestra el complemento del árbol de decisión cuando la variable *Perfil deseable* toma el valor de *no*.

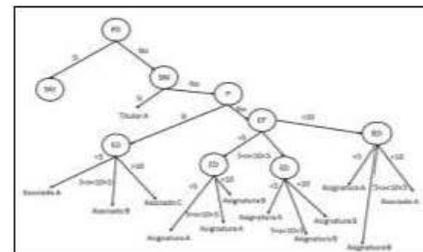


Figura 3. Sección del árbol de decisión cuando $PD = \text{no}$.



Se utilizan los 52 datos restantes para probar el modelo y se obtienen los resultados que se muestran en la Tabla 4.

Tabla 4. Datos curriculares observados.

Categoría	Elementos de Prueba	Clasificados Correctamente	Porcentaje
Asignatura A	4	4	100%
Asignatura B	10	5	50%
Asociado A	15	8	53%
Asociado B	4	3	75%
Asociado C	15	9	60%
Titular A	4	2	50%

6 CONCLUSIONES

Se puede observar que el árbol clasificador cuenta con una eficacia promedio superior al 60%, sobre todo en las categorías de Asignatura A, Asociado B y C. Para que mejore la eficacia del clasificador se debe pedir que se analicen otras características curriculares que permitan dar mayor claridad en la evaluación de prospectos de personal académico para el TESCO y seguir validando los datos que alimenten al algoritmo.

7 REFERENCIAS

[1]. Estado de México, "Gaceta de Gobierno," Mexico, 1996. Accessed: Oct. 03, 2022. [Online]. Available: https://legislacion.edomex.gob.mx/ve_periodico_oficial

[2]. C. Malagón Luque, "Clasificadores bayesianos. El algoritmo Naïve Bayes."

[3]. B. Sierra Araujo, Aprendizaje automático: conceptos básicos y avanzados, Primera., vol. 1. España: Pearson - Prentice Hall, 2006.

[4]. E. C. Vasconcelos et al., "Decision tree classifiers for star/galaxy separation," *Astronomical Journal*, vol. 141, no. 6, Jun. 2011, doi: 10.1088/0004-6256/141/6/189.

[5]. D. W. Gerdes et al., "ArborZ: Photometric redshifts using boosted decision trees," *Astrophysical Journal*, vol. 715, no. 2, pp. 823–832, 2010, doi: 10.1088/0004-637X/715/2/823.

[6]. R. E. Barrientos Martínez et al., "Árboles de decisión como herramienta en el diagnóstico médico," *Revista médica de la Universidad Veracruzana*, vol. 9, no. 2, pp. 18–24, Sep. 2009. [Online]. Available: www.uv.mx/rm

[7]. D. He, "A study on the application of decision tree algorithm in mobile marketing," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Oct. 2021. doi: 10.1088/1742-6596/2037/1/012033.

[8]. F. Syukmana et al., "Predicting Relegation Clubs in Italian Serie A with Method based C4.5 Decision Tree Algorithm," in *Journal of Physics: Conference Series*,

Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1742-6596/1471/1/012016.

[9]. O. F. Althuwaynee, A. L. Balogun, A. Aydda, and T. Gumbo, "Classification of air pollutants API Inter-Correlation using decision tree algorithms," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Feb. 2020. doi: 10.1088/1755-1315/419/1/012022.

[10]. M. Wasi, A. Sudiarto, and Fathurrahman, "Application of the Decision Tree Method to Predict Student Achievement Viewed from Final Semester Values," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jul. 2020. doi: 10.1088/1742-6596/1539/1/012027.

[11]. Y. Wang and D. Duan, "Research on risk assessment of clients before loan based on decision tree algorithm," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1774/1/012056.

J. Jones. (2007, febrero 6). *Networks* (2nd ed.) [En línea]. Disponible en: <http://www.atm.com>

[12]. N. Gangsarwijaya, R. Anindhita, and D. L. Widaningrum, "Decision tree analysis approach to determine factors that affect the quote order lead time fulfillment," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1755-1315/426/1/012122.

[13]. Y. Kustiyahningsih, B. K. Khotimah, D. R. Anamisa, M. Yusuf, T. Rahayu, and J. Purnama, "Decision Tree C 4.5 Algorithm for Classification of Poor Family Scholarship Recipients," *IOP Conf Ser Mater Sci Eng*, vol. 1125, no. 1, p. 012048, May 2021, doi: 10.1088/1757-899X/1125/1/012048.

[14]. N. R. Nurwulan and G. Selamaj, "Human daily activities recognition using decision tree," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1833/1/012039.

[15]. Tecnológico Nacional de México, "Lineamientos que establecen los requisitos y condiciones que debe reunir el personal académico que imparte planes y programas de estudio en los institutos tecnológicos descentralizados," <http://www.dgest.gob.mx/itd/academica-ditd>.

[16]. Flores Rodríguez Camilo, "Generación de árboles de decisión usando un algoritmo inspirado en la Física," *Universidad Veracruzana*, Veracruz, 2021.

[17]. J. M. Díaz-Nafria, "glossariumBITri," *Entropía o cantidad de información*, 2021.