

**Universidad Tecnológica de Panamá**

**Ingeniería en Sistemas y Computación**

**Licenciatura en ingeniería de sistemas computacionales**



**Proyecto Semestral de Probabilidad Aplicada a TIC**

**Docente:** Juan Marcos Castillo, PhD

Integrantes: Mercado, José 8-1045-2452

Ayala, Ariel 8-1037-61

Grupo: 1IL-124

## **Introducción:**

En el contexto del análisis de datos y la estadística, las bases de datos juegan un papel fundamental al almacenar grandes volúmenes de información que pueden ser estudiados mediante herramientas probabilísticas. La probabilidad permite analizar la incertidumbre inherente a muchos fenómenos del mundo real, y cuando se combina con bases de datos, se convierte en una herramienta poderosa para la toma de decisiones informadas.

El uso de bases de datos en probabilidad facilita la recopilación, organización y procesamiento de datos que se utilizan para calcular medidas estadísticas como la media, la varianza o la probabilidad de eventos. Por ejemplo, en un sistema de gestión de pacientes en un hospital, una base de datos puede almacenar información sobre enfermedades, tratamientos y resultados, lo que permite aplicar modelos probabilísticos para predecir riesgos, necesidades o comportamientos futuros.

Además, el análisis probabilístico en bases de datos también es esencial en campos como la inteligencia artificial, la minería de datos, y la ingeniería de software, donde es crucial estimar la probabilidad de eventos basados en patrones históricos. Con el crecimiento de los datos digitales, se ha vuelto cada vez más importante comprender cómo aplicar técnicas probabilísticas sobre grandes volúmenes de información de forma eficiente y confiable.

## **Justificación:**

Nosotros escogimos esta base de datos ya que los estudiantes están presentes en dicha base, el sentido de representación es notorio ya que hay bastantes estudiantes que trabajan y/o están presentes en actividades extracurriculares dentro o fuera del complejo universitario, lo que dificulta o reduce la organización o distribución correcta y adecuada del tiempo, aunque todas las personas son diferentes, hay casos similares de la misma organización del tiempo que tienen promedios completamente diferentes, gracias a esas diferencias, surgió el interés en escoger esta base de datos, con el fin de poder analizarla y representar los distintos resultados reflejados según sus respectivas variables.

Además este proyecto nos puede guiar o recomendar a organizar mejor los presentes tiempos de estudio, esto con el fin de considerar si es lo más adecuado tener actividades extracurriculares o un trabajo de medio tiempo, sin afectar los estudios universitarios; aunque según la base de datos, la carrera influye en un pequeño porcentaje al índice, ya que según los datos analizados, los estudiantes con las mismas variables solo se les vio afectado el índice por la carrera que escogieron; esto quiere decir que como tal la carrera no es que sea fácil, sino que presenta un grado de complejidad menor, lo que otorga mayor tiempo disponible sin tener que afectar el desempeño en GPA.

## **Antecedentes**

La facultad detectó un aumento en la preocupación por el bajo rendimiento y la deserción temprana.

Se planteó identificar factores académicos y personales que expliquen el GPA y ayuden a predecir quién podría necesitar apoyo.

### **Fuente de datos**

Encuesta institucional aplicada al cierre del semestre 2024-1 ( $n \approx 300$  estudiantes).

### **Variables recogidas:**

- Demográficas: Gender, Age
- Académicas: StudyHoursPerWeek, AttendanceRate, GPA, Major
- Situacionales: PartTimeJob (Sí/No), ExtraCurricularActivities (Sí/No)

## Definición del problema

La incógnita que formulamos con los datos obtenidos fue: Las horas de estudio, asistencias, trabajos y/o actividades extracurriculares afectan en nuestro desempeño académico?

Esta incógnita la formulamos ya que eran las variables presentes en la base de datos, aunque también tenemos la edad; esta variable representa un cambio mínimo entre los diferentes índices, así que la interrogante principal es la anterior mencionada.

Al analizar las variables en nuestra base de datos; nos dimos cuenta de que, a mayor tiempo de estudio disponible, asistencia casi perfecta y no tener responsabilidades laborales; influyen positivamente en el índice mostrando que la mayoría de estos estudiantes presentan un índice arriba de 2.75.

Por el contrario, si se presenta muy pocas veces y no le dedica mucho tiempo de estudio, ya sea por trabajo o actividades fuera del periodo universitario, el índice decrece significativamente.

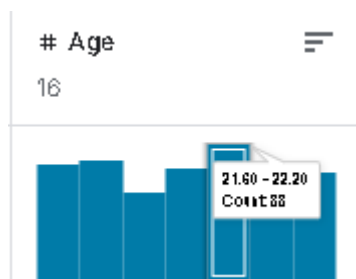
Aunque quiero resaltar que hay estudiantes con las horas de asistencia y de estudio al mínimo, presentan un índice elevado y viceversa; esto lo conocemos como puntos atípicos, es decir valores que no coinciden con el orden o patrones de los demás.

## Avance de Estadísticos

Las medidas de tendencia central:

Promedio de edad de las mujeres		20.77359263	
Promedio de edad de los hombres		21.14675913	

La moda de edad es de 22 años con 88 estudiantes



La bimodalidad de las horas de estudio son los siguientes intervalos: 1-2.9 y 25.7-27.6

Clase	Frecuencia
1.0 - 2.9	30
2.9 - 4.8	27
4.8 - 6.7	30
6.7 - 8.6	24
8.6 - 10.5	28
10.5 - 12.4	23
12.4 - 14.3	25
14.3 - 16.2	16
16.2 - 18.1	26
18.1 - 20.0	10
20.0 - 21.9	23
21.9 - 23.8	29
23.8 - 25.7	23
25.7 - 27.6	30
27.6 - 29.5	26
29.5 - 31.4	27
31.4 - 33.3	27
33.3 - 35.2	24
35.2 - 37.1	29
37.1 - 39.0	23

La moda del GPA es del rango de 2.3 a 2.4

Clase	Frecuencia
2.0 - 2.1	19
2.1 - 2.2	23
2.2 - 2.3	26
2.3 - 2.4	37
2.4 - 2.5	24
2.5 - 2.6	25
2.6 - 2.7	22
2.7 - 2.8	29
2.8 - 2.9	22
2.9 - 3.0	20
3.0 - 3.09	28
3.09 - 3.19	25
3.19 - 3.29	34
3.29 - 3.39	28
3.39 - 3.49	17
3.49 - 3.59	29
3.59 - 3.69	27
3.69 - 3.79	24
3.79 - 3.89	19
3.89 - 3.99	22

Resultados	Valores
Promedio del GPA predictivo:	2.15521484
Porcentaje de error promedio %	28.83515247
Promedio del GPA real	2.98516
Deupuración	27.80236771
Formula propuesta para la predicción del GPA a base de los patrones anteriores	



## Visualización de Gráfico

