

Pontificia Universidad Católica del Ecuador
Facultad De Ingeniería



TEMA:

ANÁLISIS COMPARATIVO DE ALGORITMOS DE MACHINE LEARNING APLICADOS EN LA
GESTIÓN DE RIESGOS CUANTITATIVOS EN PROYECTOS PREDICTIVOS Y ADAPTATIVOS

AUTOR:

FRANCISCO MIGUEL OSEJO DOMÍNGUEZ

TRABAJO PREVIA A LA OBTENCIÓN DEL TÍTULO DE MAGISTER EN SISTEMAS DE
INFORMACIÓN MENCIÓN DATA SCIENCE

QUITO, agosto - 2023

DEDICATORIA

Con mucho amor dedico este proyecto de investigación a mis hijos Gabriel Alejandro, Ricardo André y en especial a mi amada esposa Jane Elizabeth por haber confiado en mí, darme todo su apoyo incondicional y comprensión.

Es un ejemplo de superación, esfuerzo, dedicación y disciplina.

Amados hijos, todo es posible en la vida, nunca se rindan, persigan sus objetivos.

AGRADECIMIENTO

Agradezco a Dios por la oportunidad de vivir, experimentar y soñar, a mi preciosa madre que me enseñó a luchar y a nunca rendirme ante las vicisitudes de la vida.

Mi mayor agradecimiento a mi hermano Edgar y a mi padre Guillermo por haber financiado esta aventura, a mis suegros Marco y Rosa Adriana de La Cruz por el constante apoyo, a Oswaldo Espinosa mi asesor académico por la guía y tiempo otorgado.

A mis hermanas por la motivación y empuje emocional.

Y finalmente a la vida por permitirme haber llegado hasta este momento.

RESUMEN

El presente trabajo tiene como fin, realizar el análisis comparativo de la aplicación de algoritmos de Machine Learning en la gestión de riesgos cuantitativos en proyectos predictivos y adaptativos.

Los algoritmos seleccionados para el desarrollo de los modelos fueron: Simulación de Monte Carlo aplicando Distribución Triangular y Beta – Pert, Regresión Lineal Simple, Regresión Lineal Múltiple y Arboles de Decisión.

Se utilizó la metodología CRISP-DM para sistematizar la información y el flujo de actividades para el desarrollo y posterior evaluación de los modelos analizados.

Los modelos se desarrollaron con Python, los datasets seleccionados son relacionados a cronograma y presupuesto de la automatización del sistema BAS y seguridad electrónica de un proyecto del sector de la construcción.

Se detalla los análisis e interpretación de los resultados y coeficientes obtenidos en el desarrollo de cada uno de los algoritmos aplicados.

Se obtiene un análisis comparativo de la aplicación de los algoritmos seleccionados, modelos y resultados obtenidos, así como las conclusiones y recomendaciones pertinentes.

ÍNDICE

ÍNDICE DE GRÁFICOS	VII
ÍNDICE DE TABLAS	XIV
1 INTRODUCCIÓN	1
1.1 Antecedentes	1
1.2 Planteamiento del problema.....	2
1.3 Justificación	3
1.4 Objetivos	3
1.4.1 Objetivo General	3
1.4.2 Objetivos Específicos	3
2 MARCO TEÓRICO Y CONCEPTUAL	5
2.1 La gestión de riesgos en proyectos	5
2.1.1 ¿Qué es un riesgo?	6
2.1.2 Riesgos Negativos.....	6
2.1.3 Riesgos Positivos.....	6
2.1.4 Categorías de riesgos.....	7
2.1.5 Probabilidad e impactos de los riesgos.	8
2.1.6 Matriz de probabilidad e impacto.....	9
2.1.7 Análisis Cualitativo de riesgos	10
2.1.8 Análisis Cuantitativo de Riesgos	11
2.2 Línea Base.....	12
2.2.1 Línea Base del Alcance	12
2.2.2 Línea Base del Cronograma	13
2.2.3 Línea Base del Costo.....	13
2.3 Estimaciones.....	13
2.3.1 Estimación Análoga	13
2.3.2 Estimación Paramétrica.....	14
2.3.3 Estimación Basada En Tres Valores o PERT.....	15
2.4 Distribuciones Probabilísticas.....	16
2.4.1 Distribuciones Probabilísticas Para Cuantificar los Riesgos.....	17
2.4.2 Distribución Normal	17
2.4.3 Distribución Triangular	18
2.4.4 Distribución Uniforme	19

2.4.5	Distribución Beta o Pert	20
2.5	Lógica Difusa.....	20
2.5.1	Lógica Difusa en Inteligencia Artificial.....	20
2.6	Machine Learning.....	21
2.6.1	Aprendizaje Supervisado	21
2.6.2	Aprendizaje No Supervisado	22
2.6.3	Modelo	22
2.7	Algoritmos de machine learning relevantes que se aplican en la gestión de riesgos cuantitativos.....	23
2.7.1	Estudio McKinsey, “Notes from the AI frontier insights from hundreds of use cases.”	23
2.7.2	Herramientas para análisis numérico de riesgos	24
2.7.3	Algoritmos de machine learning seleccionados para aplicar en la gestión de riesgos cuantitativos.....	25
2.8	Monte Carlo.....	25
2.8.1	¿Qué es la simulación Monte Carlo?	25
2.8.2	¿Cómo funciona la simulación Monte Carlo?	26
2.8.3	Cómo utilizar los métodos de Monte Carlo.....	27
2.9	Arboles de decisión	28
2.10	Análisis de regresión.....	30
2.11	Librería SciPy	30
2.12	Estadísticas (SciPy.stats)	31
2.13	Metodología CRISP-DM	31
3	Metodología y Técnicas	33
3.1	Caso de uso	33
3.2	Comprensión del negocio.....	33
3.2.1	Contrato de Concesión y Alianza Estratégica.	33
3.2.2	Determinar Objetivos del Negocio	35
3.2.3	Criterios de éxito del negocio.....	36
3.2.4	Evaluar la situación.....	36
3.2.5	Inventario de Recursos	36
3.2.6	Requisitos, supuestos y restricciones.....	37
3.2.7	Riesgos y contingencias	37
3.2.8	Costos	38
3.2.9	Determinación de los objetivos de modelación.....	38
3.2.10	Plan de proyecto.....	39
3.3	Comprensión de datos	39

3.3.1	Recopilación de datos iniciales.....	39
3.3.2	Describiendo los datos	40
3.3.3	Exploración de los datos.....	42
3.4	Preparación de datos.....	43
3.4.1	Cronograma – tiempo	44
3.4.2	Presupuesto – costo	45
3.4.3	Selección de datos.....	45
3.4.4	Cronograma – tiempo	46
3.4.5	Presupuesto – costo	48
3.4.6	Limpieza de datos.....	48
3.4.7	Construcción de nuevos datos	49
3.5	Modelado	51
3.5.1	Selección de técnicas de modelado	52
3.5.2	Construyendo los modelos.....	52
3.6	Evaluación.....	120
3.6.1	Evaluación de Resultados	120
4	Análisis Comparativo de Modelos Aplicados	127
4.1	Simulación de Monte Carlo	127
4.1.1	Simulación de Monte Carlo al Dataset de Cronograma – Tiempo	127
4.1.2	Simulación de Montecarlo al Dataset de Presupuesto – Costo.....	129
4.2	Resumen del análisis comparativo de los modelos aplicados.....	133
5	CONCLUSIONES Y RECOMENDACIONES	135
5.1	Conclusiones.....	135
5.2	Recomendaciones	137
	BIBLIOGRAFÍA.....	138

ÍNDICE DE GRÁFICOS

Gráfico 1 Estadísticas de proyectos de inversión según EY. Autor: (Young, 2009)	1
Gráfico 2 Representa los procesos de la gestión de riesgos Autor: (Buchtik, 2015)	5
Gráfico 3 Tipos de riesgos Autor: (Buchtik, 2015)	7
Gráfico 4 Extracto de una estructura de desglose de los riesgos (RBS) Autor: (PMBOK_V6, 2017)	8
Gráfico 5 Matriz de probabilidad e impacto con esquema de puntuación. Autor: (PMBOK_V6, 2017)	10
Gráfico 6 Tipos de análisis de riesgos. Autor: (Buchtik, 2015)	12
Gráfico 7 Distribuciones probabilísticas usadas en el análisis numérico de riesgos. Autor: (Buchtic, 2023).....	16
Gráfico 8 Distribuciones probabilísticas normal Autor: (Buchtic, 2023)	17
Gráfico 9 Ej. Distribuciones normales con distinto riesgo Autor: (Buchtic, 2023).....	18
Gráfico 10 Distribuciones triangulares. Autor: (Buchtic, 2023).....	19
Gráfico 11 Distribuciones uniforme. Autor: (Buchtic, 2023)	19
Gráfico 12 Distribución beta Autor: (Buchtic, 2023)	20
Gráfico 13 Mapa de calor: Relevancia de la técnica para las funciones Autor: (McKinsey&Company, 2018)	24
Gráfico 14 Caja de herramientas para analizar los riesgos numéricamente Autor: (Buchtik L. , 2015)	25
Gráfico 15 Ejemplo de árbol de decisiones de pesca Autor: (Brian Boucheron & Lisa Tagliaferri, 2022).....	29
Gráfico 16 El ciclo de vida de la minería de datos Autor: (CRISP-DM, 2021)	32

Gráfico 17 Cronograma – tiempo Autor: (Osejo Francisco, 2023)	43
Gráfico 18 Presupuesto – costo Autor: (Osejo Francisco, 2023)	43
Gráfico 19 Cronograma – tiempo en formato Excel Autor: (Osejo Francisco, 2023)	44
Gráfico 20 Presupuesto – costo en formato Excel Autor: (Osejo Francisco, 2023)	45
Gráfico 21 FID – Monitor informativo arribos y salidas Autor: (Osejo Francisco, 2023)	47
Gráfico 22 FID – monitor curvo informativo arribos y salidas Autor: (Osejo Francisco, 2023).....	47
Gráfico 23 Cronograma – tiempo con nuevas variables Autor: (Osejo Francisco, 2023)50	
Gráfico 24 Presupuesto – costo con nuevas variables Autor: (Osejo Francisco, 2023)..	51
Gráfico 25 Simulación monte carlo – contenido dataset cronograma – tiempo Autor: (Osejo Francisco, 2023).....	53
Gráfico 26 Simulación monte carlo - código en Python distribución probabilística Triangular Cronograma – tiempo Autor: (Osejo Francisco, 2023).....	54
Gráfico 27 Simulación monte carlo - código en Python distribución probabilística triangular cronograma – tiempo Autor: (Osejo Francisco, 2023).....	55
Gráfico 28 Simulación monte carlo - código en Python distribución probabilística triangular cronograma – tiempo Autor: (Osejo Francisco, 2023).....	56
Gráfico 29 Histograma distribución triangular – aplicando monte carlo (diez mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	57
Gráfico 30 Histograma distribución triangular – aplicando monte carlo (diez mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	58
Gráfico 31 Histograma distribución triangular – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	60

Gráfico 32 Histograma distribución triangular – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	61
Gráfico 33 Histograma distribución triangular – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	62
Gráfico 34 Simulación monte carlo - código en Python distribución beta o pert cronograma – tiempo Autor: (Osejo Francisco, 2023)	65
Gráfico 35 Simulación monte carlo - código en Python distribución beta o pert cronograma – tiempo Autor: (Osejo Francisco, 2023)	66
Gráfico 36 Simulación monte carlo - código en Python distribución beta o pert cronograma – tiempo Autor: (Osejo Francisco, 2023)	67
Gráfico 37 Simulación monte carlo - código en Python distribución beta o pert cronograma – tiempo Autor: (Osejo Francisco, 2023)	67
Gráfico 38 Histograma distribución beta o pert – aplicando monte carlo (diez mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	68
Gráfico 39 Histograma distribución beta o pert – aplicando monte carlo (diez mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	69
Gráfico 40 Histograma distribución beta o pert – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	71
Gráfico 41 Histograma distribución beta o pert – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	72
Gráfico 42 Histograma distribución beta o pert – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023).....	73
Gráfico 43 Simulación monte carlo – contenido dataset presupuesto – costo Autor: (Osejo Francisco, 2023).....	76

Gráfico 44 Simulación monte carlo - código en Python distribución probabilística triangular presupuesto – costo Autor: (Osejo Francisco, 2023)	77
Gráfico 45 Simulación monte carlo - código en Python distribución probabilística triangular presupuesto – costo Autor: (Osejo Francisco, 2023)	78
Gráfico 46 Simulación monte carlo - código en Python distribución probabilística triangular presupuesto – costo Autor: (Osejo Francisco, 2023)	78
Gráfico 47 Simulación monte carlo - código en Python distribución probabilística triangular presupuesto – costo Autor: (Osejo Francisco, 2023)	79
Gráfico 48 Histograma Distribución Triangular – Aplicando Monte Carlo (Diez Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	80
Gráfico 49 Histograma Distribución Triangular – Aplicando Monte Carlo (Diez Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	80
Gráfico 50 Histograma Distribución Triangular – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	83
Gráfico 51 Histograma Distribución Triangular – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	83
Gráfico 52 Histograma Distribución Triangular – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	84
Gráfico 53 Simulación Monte Carlo - Código en Python Distribución Beta o Pert Presupuesto – Costo Autor: (Osejo Francisco, 2023)	88
Gráfico 54 Simulación Monte Carlo - Código en Python Distribución Beta o Pert Presupuesto – Costo Autor: (Osejo Francisco, 2023)	89
Gráfico 55 Simulación Monte Carlo - Código en Python Distribución Beta o Pert Presupuesto – Costo Autor: (Osejo Francisco, 2023)	89

Gráfico 56 Simulación Monte Carlo - Código en Python Distribución Beta o Pert	
Presupuesto – Costo Autor: (Osejo Francisco, 2023)	90
Gráfico 57 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Diez Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	91
Gráfico 58 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Diez Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	91
Gráfico 59 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	94
Gráfico 60 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	94
Gráfico 61 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)	95
Gráfico 62 Regresión Lineal Simple Dataset Presupuesto – Costo Autor: (Osejo Francisco, 2023)	99
Gráfico 63 Regresión Lineal Simple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	100
Gráfico 64 Regresión Lineal Simple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	101
Gráfico 65 Regresión Lineal Simple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	101
Gráfico 66 Regresión Lineal Simple – Gráfico de dispersión Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	102
Gráfico 67 Regresión Lineal Simple – Datos y_test – Costo Autor: (Osejo Francisco, 2023).....	103

Gráfico 68 Regresión Lineal Simple – Datos y_predit – Costo Autor: (Osejo Francisco, 2023).....	103
Gráfico 69 Regresión Lineal Simple – Gráfico de dispersión con la línea de regresión – Costo Autor: (Osejo Francisco, 2023)	104
Gráfico 70 Regresión Lineal Simple – Datos del modelo - Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	105
Gráfico 71 Regresión Lineal Múltiple Dataset Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	106
Gráfico 72 Regresión Lineal Múltiple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	107
Gráfico 73 Regresión Lineal Múltiple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	108
Gráfico 74 Regresión Lineal Múltiple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	108
Gráfico 75 Regresión Lineal Múltiple – Datos y_test – Costo Autor: (Osejo Francisco, 2023).....	109
Gráfico 76 Regresión Lineal Múltiple – Datos y_pred – Costo Autor: (Osejo Francisco, 2023).....	110
Gráfico 77 Regresión Lineal Múltiple – Gráfico de dispersión con la línea de regresión – Costo Autor: (Osejo Francisco, 2023)	110
Gráfico 78 Regresión Lineal Múltiple – Datos del modelo - Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	111
Gráfico 79 Arboles de Decisión Dataset Presupuesto – Costo Autor: (Osejo Francisco, 2023).....	113

Gráfico 80 Arboles de Decisión - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)	113
Gráfico 81 Arboles de Decisión - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)	114
Gráfico 82 Arboles de Decisión - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)	114
Gráfico 83 Arboles de Decisión – Datos y_test – Costo Autor: (Osejo Francisco, 2023)	115
Gráfico 84 Arboles de Decisión – Datos y_pred – Costo Autor: (Osejo Francisco, 2023)	115
Gráfico 85 Arboles de Decisión – Gráfico de dispersión – Costo Autor : (Osejo Francisco, 2023)	116
Gráfico 86 Arboles de Decisión – Predicción – Costo Autor: (Osejo Francisco, 2023)	116
Gráfico 87 Arboles de Decisión – Precisión del Modelo – Costo Autor: (Osejo Francisco, 2023)	117
Gráfico 88 Arboles de Decisión con porcentaje de frecuencia - Presupuesto – Costo Autor: (Osejo Francisco, 2023)	118
Gráfico 89 Arboles de Decisión con frecuencia - Presupuesto – Costo Autor: (Osejo Francisco, 2023)	120
Gráfico 90 Histograma Distribución Beta - Cronograma – Tiempo Autor: (Osejo Francisco, 2023)	121
Gráfico 91 Histograma Distribución Beta - Presupuesto – Costo Autor: (Osejo Francisco, 2023)	123

Gráfico 92 Arboles de Decisión con frecuencia - Presupuesto – Costo Autor: (Osejo Francisco, 2023)	125
--	-----

ÍNDICE DE TABLAS

Tabla 1 Ejemplo de definiciones para probabilidad e impacto. Autor: (PMBOK_V6, 2017)	9
Tabla 2 Riesgos y contingencias del proyecto. Autor: (Osejo Francisco, 2023)	37
Tabla 3 Costos proyecto. Autor: (Osejo Francisco, 2023)	38
Tabla 4 Plan del proyecto. Autor: (Osejo Francisco, 2023)	39
Tabla 5 Estructura de datos de dataset cronograma tiempo. Autor: (Osejo Francisco, 2023)	41
Tabla 6 Estructura de datos de dataset presupuesto costo. Autor: (Osejo Francisco, 2023)	42
Tabla 7 Datos seleccionados dataset cronograma - tiempo. Autor: (Osejo Francisco, 2023)	48
Tabla 8 Datos seleccionados dataset presupuesto - costo. Autor: (Osejo Francisco, 2023)	48
Tabla 9 Nuevas variables dataset cronograma - tiempo. Autor: (Osejo Francisco, 2023)	50
Tabla 10 Nuevas variables dataset presupuesto - costo. Autor: (Osejo Francisco, 2023)	51
Tabla 11 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000 y 10,000 Autor: (Osejo Francisco, 2023)	58

Tabla 12 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000, 10,000 y 20,000 Autor: (Osejo Francisco, 2023)	61
Tabla 13 Cuadro comparativo de información levantada en cada iteración aplicando monte carlo conjuntamente con la distribución triangular Autor: (Osejo Francisco, 2023).....	62
Tabla 14 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000 y 10,000 Autor: (Osejo Francisco, 2023).....	69
Tabla 15 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000, 10,000 y 20,000 Autor: (Osejo Francisco, 2023)	72
Tabla 16 Cuadro comparativo de información levantada en cada iteración aplicando monte carlo conjuntamente con la distribución beta o pert Autor: (Osejo Francisco, 2023).....	73
Tabla 17 Costo en millones de dólares de norte américa obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000 y 10,000 Autor: (Osejo Francisco, 2023) 81	
Tabla 18 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000, 10,000 y 20,000 Autor: (Osejo Francisco, 2023)	84
Tabla 19 Cuadro comparativo de información levantada en cada iteración aplicando monte carlo conjuntamente con la distribución triangular Autor: (Osejo Francisco, 2023).....	85
Tabla 20 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000 y 10,000 Autor: (Osejo Francisco, 2023).....	92
Tabla 21 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000, 10,000 y 20,000 Autor: (Osejo Francisco, 2023)	95

Tabla 22 Cuadro comparativo de información levantada en cada iteración aplicando monte carlo conjuntamente con la distribución beta o pert Autor: (Osejo Francisco, 2023).....	96
Tabla 23 Análisis estimación más probable original Autor: (Osejo Francisco, 2023) ...	121
Tabla 24 Análisis estimación más probable original - presupuesto – costo Autor: (Osejo Francisco, 2023).....	123
Tabla 25 Actividades ramal B - árbol de decisión - presupuesto – costo Autor: (Osejo Francisco, 2023).....	126
Tabla 26 Análisis comparativo de modelos aplicados. Autor: (Osejo Francisco, 2023)	133
Tabla 27 Cuadro comparativo simulación monte carlo – dataset cronograma – tiempo. Autor: (Osejo Francisco, 2023)	127
Tabla 28 Cuadro comparativo simulación monte carlo – dataset presupuesto – Costo. Autor: (Osejo Francisco, 2023)	130

CAPÍTULO I: INTRODUCCIÓN

1 INTRODUCCIÓN

1.1 Antecedentes

La gestión de riesgos en proyectos es un factor clave para el éxito de la entrega de un proyecto.

En base a un estudio de Ernst & Young que demuestra que la mayor parte de los grandes proyectos de inversión no alcanzan sus objetivos, termina tarde o con importantes sobrecostos (Young, 2009).

Gráfico 1 Estadísticas de proyectos de inversión según EY. Autor: (Young, 2009)

- 50% de los proyectos sobreasan el presupuesto inicial.
- 58% de ellos se entregan tarde.
- 42% tienen algún defecto luego de terminarse.
- 30% a 40% del costo del proyecto deriva en retrabajos.
- 59% de los gerentes generales o financieros reconocen que no siguen un proceso estructurado de gestión de riesgos de proyectos
- Muchas empresas tienen al menos 10 grupos de personas trabajando independientemente en diferentes actividades de riesgos.
- Más del 70% de los proyectos no logra los objetivos esperados.

Según un estudio realizado en 35 firmas de construcción en Estados Unidos con proyectos de más de cien millones de dólares, el **84% de los proyectos de infraestructura no terminan a tiempo y el 86% termina sobre el presupuesto**. Además, el 82% tiene hasta 20% de órdenes de cambio, el 11% tiene entre 21% y 50% de órdenes y el 7% tiene más de 61% de órdenes de cambio (MacGraw-Hill, 2011).

Los riesgos más frecuentes en los proyectos de construcción, infraestructura, tecnología y muchos de ellos están bajo el control de las compañías ... reforzando el hecho de que las estrategias de respuesta a los riesgos pueden tener un impacto positivo (Buchtik, 2015).

Según una investigación realizada por el PMI (Project Management Institute), el 88% de las organizaciones de alto rendimiento a menudo usan técnicas de gestión de riesgos, comparado con el 54% de las organizaciones de bajo desempeño Podría llevar a pensar que para las organizaciones más sobresalientes es muy importante aplicar gestión de

riesgos. Factores más críticos de éxito en proyectos podría ser reconocer y cuantificar los riesgos (Pulse of the Profession Research, 2010).

1.2 Planteamiento del problema

Dentro de la Gestión de Proyectos tanto predictivos como adaptativos e híbridos, el análisis de gestión de riesgos cumple un papel vital dentro de la planificación de un proyecto predictivo, así como en la determinación de un Producto Mínimo Viable (VMP) en proyectos adaptativos, y más aún en proyectos híbridos.

Dentro de la Gestión del Riesgo existen dos grupos: riesgos cualitativos y riesgos cuantitativos.

Los riesgos cualitativos son generalmente tratados y gestionados apropiadamente, ya que no existe mayor problemática.

Los riesgos cuantitativos por lo general son identificados inicialmente y poco desarrollados posteriormente, ya que, existe cierta complejidad tanto técnica como económica.

Lamentablemente en nuestro medio, la gestión de proyectos en su gran mayoría de veces se los gestiona desde una visión empírica, dejando de lado, áreas muy importantes como la gestión de riesgos, entre otras.

El éxito de un proyecto se basa en la determinación de la Línea Base (Alcance, Cronograma, Costo) del proyecto, y evidentemente en la ejecución, así como al monitoreo y control. Esto aplica a proyectos predictivos, adaptativos e híbridos.

Las complicaciones de los proyectos se dan en la fase de ejecución, en donde se presentan los problemas reales, circunstancias ocasionadas por la falta de un análisis más técnico de la línea base. La piedra angular de la línea base son las estimaciones de tiempo, costo, recursos, etc. Sustentadas en un correcto análisis de riesgos, es decir, sin una correcta gestión de riesgos, la línea base no tiene sustento. Ya que su mínima variación y al no tener un plan de respuesta de riesgos, el proyecto está destinado al fracaso.

Por lo general, la Gestión de Riesgos no se la realiza, en muchos de los casos, y en los pocos casos no son bien gestionados.

La Gestión de Riesgos Cuantitativos son de relevancia y de vital importancia, por lo tanto, para su determinación se usan modelos matemáticos y estadísticos, así como software especializado, logrando obtener cierta precisión.

La presente formulación de Trabajo de Titulación plantea la siguiente pregunta:

¿Qué algoritmos de Machine Learning pueden ser aplicados en la Gestión de Riesgos Cuantitativos en Proyectos Predictivos y Adaptativos?

1.3 Justificación

La Gestión de Riesgos Cuantitativos son de relevancia y de vital importancia, por lo tanto, para su determinación se usan modelos matemáticos y estadísticos, así como software especializado, logrando obtener cierta precisión.

El poder determinar la aplicación de algoritmos de Machine Learning en el análisis de Riesgos Cuantitativos nos permitirá obtener niveles de precisión en las estimaciones, predicciones, tanto en costo como en tiempo. Pudiendo comparar entre ellos y obtener las conclusiones del estudio.

1.4 Objetivos

1.4.1 Objetivo General

Identificar los algoritmos de Machine Learning más relevantes que se apliquen a la gestión de riesgos cuantitativos en proyectos predictivos y adaptativos, realizando posteriormente un análisis comparativo.

1.4.2 Objetivos Específicos

- Determinar diferencias entre riesgos cualitativos vs riesgos cuantitativos.
- Identificar y seleccionar los algoritmos de machine learning relevantes que se apliquen en la gestión de riesgos cuantitativos en proyectos predictivos y adaptativos.
- Seleccionar el Dataset que permita recrear la aplicación de los algoritmos de machine learning seleccionados.
- Utilizar la metodología CRISP DM para demostrar el uso de los algoritmos de machine learning seleccionados, conjuntamente con el dataset y su aplicación en la gestión de riesgos cuantitativos en proyectos predictivos y adaptativos.

- Determinar el análisis comparativo y conclusiones de la aplicación de algoritmos de machine Learning seleccionados en la aplicación de la gestión de riesgos cuantitativos en proyectos predictivos y adaptativos.

CAPÍTULO II: FUNDAMENTACIÓN TEÓRICA

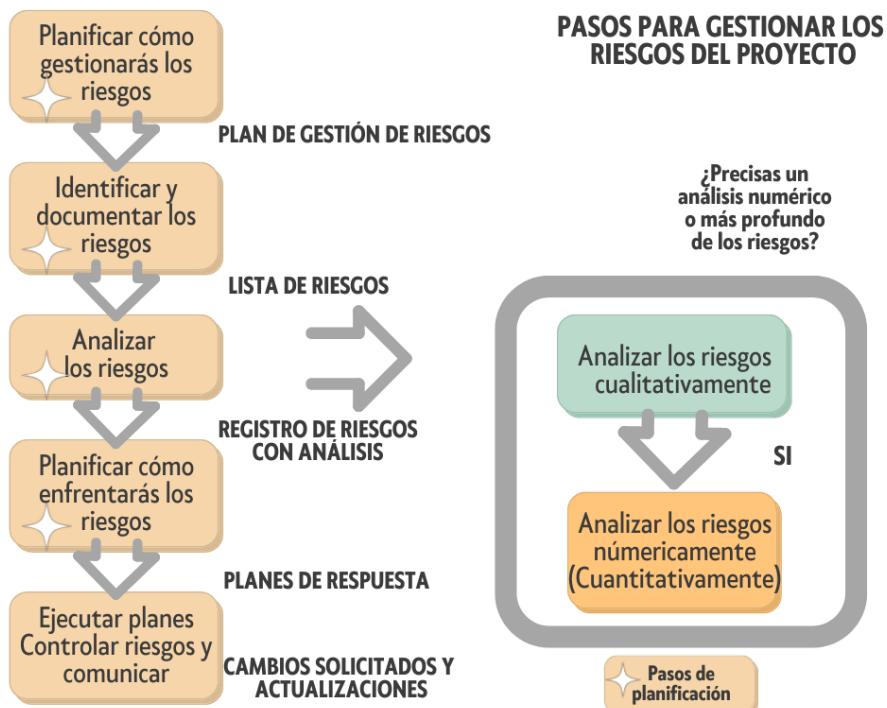
2 MARCO TEÓRICO Y CONCEPTUAL

2.1 La gestión de riesgos en proyectos

La gestión de riesgos en proyectos es tratar con los riesgos antes de que se vuelvan problemas. Es preocuparse de ser proactivo en vez de reactivo. Incluye planificar la forma en que se va a gestionar los riesgos, identificar, documentar, y analizar los riesgos, planificar cómo enfrentarlos, implementar los planes, y luego controlarlos. Busca aumentar la probabilidad y el impacto de los eventos positivos, y disminuir la probabilidad y el impacto de los eventos adversos al proyecto (PMBOK_V7, 2021).

En un lenguaje común, gestionar los riesgos es dejar de ser bombero que vive apagando incendios, y en su lugar, planificar qué hacer para evitar los incendios, o por lo menos para minimizarlos (Buchtik, 2015).

Gráfico 2 Representa los procesos de la gestión de riesgos Autor: (Buchtik, 2015)



2.1.1 ¿Qué es un riesgo?

Riesgo es el efecto de la incertidumbre sobre los objetivos (ISO 31000, 2009).

Todos los proyectos son riesgosos, ya que son emprendimientos únicos con diferentes grados de complejidad que tienen como objetivo ofrecer beneficios. Se dedican a esto dentro de un contexto de restricciones y suposiciones al tiempo que responden a las expectativas de los interesados, las que pueden ser contradictorias y cambiantes. Las organizaciones deben elegir enfrentar el riesgo del proyecto de una manera controlada e intencional para crear valor equilibrado al mismo tiempo el riesgo y la recompensa (PMBOK_V6, 2017).

Un riesgo no es un problema, un problema ya ocurrió. Un riesgo podría ocurrir. Gestionar los riesgos es gestionar problemas potenciales. Si la economía ya se devaluó no hay un riesgo sino un problema a tratar. Los directores de proyectos exitosos no se enfocan en los problemas sino en prevenirlos. La gestión proactiva ahorra tiempo y dinero, y minimiza las incertidumbres (Buchtik, 2015).

2.1.2 Riesgos Negativos

En general es más fácil identificar riesgos negativos que positivos. Hay más cosas que pueden salir mal de las que pueden salir bien. La mayoría de los riesgos hacen que el proyecto demore más y sea más caro; y no viceversa (Buchtik, 2015)

2.1.3 Riesgos Positivos

Históricamente la palabra riesgo se asoció a lo malo, a las pérdidas y amenazas. Sin embargo, existe el concepto de riesgo positivo, el cual aborda las oportunidades. Pensar en riesgos positivos requiere de mayor creatividad. Implica pensar en oportunidades de reducir costos o duraciones de tareas, de simplificar, de obtener bonos o ganancias según como se trabaje, etc. Para identificar este tipo de riesgos, me resulta más fácil usar la palabra oportunidad en lugar de usar riesgo positivo (Buchtik, 2015)

Gráfico 3 Tipos de riesgos Autor: (Buchtik, 2015)

	TIPOS DE RIESGOS	
	NEGATIVOS	POSITIVOS
Llamado	Amenazas	Oportunidad
Definición	Situación que si ocurre, puede impactar negativamente algún objeto del proyecto.	Situación que si ocurre puede impactar positivamente algún objetivo del proyecto.
Ejemplo	Si planificas un concierto, el riesgo es que el cantante por algún motivo no pueda asistir.	Si planificas un concierto, existe la oportunidad de doblar la asistencia del año previo.
Causa * (1 o +)	Se demoró el vuelo del cantante o el cantante se enfermo.	Se realiza el concierto trayendo al cantante más reconocido del momento.
Impacto si ocurre (1 o +)	Se cancela el concierto.	Se generarían \$ 400.000 más que el año anterior.
¿Qué haces?	Minimizarlos.	Maximizarlos.

2.1.4 Categorías de riesgos

Proporciona un medio para agrupar los riesgos individuales de cada proyecto. Una forma común de estructurar las categorías de riesgo es por medio de una estructura de desglose de los riesgos (RBS), que es una representación jerárquica de las posibles fuentes de riesgos. Una RBS ayuda al equipo del proyecto a tener en cuenta toda la gama de fuentes a partir de las cuales pueden derivarse los riesgos individuales del proyecto. Esto puede ser útil en la identificación de riesgos o al categorizar riesgos identificados. Cuando no se utiliza una RBS, una organización puede utilizar un marco personalizado de categorización de riesgos, que puede adoptar la forma de una simple lista de categorías o de una estructura basada en los objetivos del proyecto (PMBOK_V6, 2017).

Gráfico 4 Extracto de una estructura de desglose de los riesgos (RBS) Autor: (PMBOK_V6, 2017)

NIVEL 0 de RBS	NIVEL 1 de RBS	MNIVEL 2 de RBS
0. TODAS LAS FUENTES DE RIESGO DEL PROYECTO	1. RIESGO TÉCNICO	1.1 Definición del alcance
		1.2 Definición de los requisitos
		1.3 Estimaciones, supuestos y restricciones
		1.4 Procesos técnicos
		1.5 Tecnología
	2. RIESGO DE GESTIÓN	1.6 Etc.
		2.1 Dirección de proyectos
		2.2 Dirección del programa / portafolio
		2.3 Gestión de las operaciones
		2.4 Organización
		2.5 Comunicación
	3. RIESGO COMERCIAL	2.6 Etc.
		3.1 Términos y condiciones contractuales
		3.2 Contratación interna
		3.3 Proveedores y vendedores
		3.4 Subcontratos
		3.5 Estabilidad de los clientes
		3.6 Asociación y empresas conjuntas
	4. RIESGO EXTERNO	3.7 ETC
		4.1 Legislación
		4.2 Tasas de cambio
		4.3 Sitios/Instalaciones
		4.4 Ambiental/clima
		4.5 Etc.

2.1.5 Probabilidad e impactos de los riesgos.

Las definiciones de la probabilidad e impacto de los riesgos son específicas al contexto del proyecto y reflejan el apetito al riesgo y los umbrales de la organización y los interesados clave. El proyecto puede generar definiciones específicas de los niveles de probabilidad e impacto, o puede comenzar con definiciones generales proporcionadas por la organización. El número de niveles refleja el grado de detalle requerido para el proceso Gestión de Riesgos del Proyecto, utilizando más niveles para un enfoque más detallado del riesgo (típicamente cinco niveles), y menos para un proceso sencillo (normalmente tres). Estas escalas se pueden utilizar para evaluar las amenazas y las

oportunidades mediante la interpretación de las definiciones de impacto como negativo para las amenazas (retardo, costo adicional y déficit de desempeño) y positivo para las oportunidades (reducción de tiempo o del costo y mejora del desempeño) (PMBOK_V6, 2017).

Tabla 1 Ejemplo de definiciones para probabilidad e impacto. Autor: (PMBOK_V6, 2017)

ESCALA	PROBABILIDAD	(+/-) IMPACTO SOBRE LOS OBJETOS DEL PROYECTO		
		TIEMPO	COSTO	CALIDAD
Muy alto	> 70%	> 6 meses	>\$5M	Impacto muy significativo sobre la funcionalidad general
Alto	51 - 70%	3-6 meses	\$1M - \$5M	Impacto significativo sobre la funcionalidad general
Mediano	31 - 50%	1-3 meses	\$501K - \$1M	Algún impacto sobre áreas funcionales clave
Bajo	11 - 30%	1-4 meses	\$100K- \$500K	Impacto menor sobre la funcionalidad general
Muy Bajo	1 - 10%	1 semana	<\$100K	Impacto menor sobre las funciones secundarias
Nulo	< 1%	Sin cambio	Sin cambio	Ningún cambio en la funcionalidad

2.1.6 Matriz de probabilidad e impacto.

Las reglas de priorización pueden ser específicas por la organización con anterioridad al proyecto y ser incluidas en los activos de los procesos de la organización, o pueden ser adaptadas para el proyecto específico. Las oportunidades y las amenazas están representadas en una matriz común de probabilidad e impacto utilizando definiciones de impacto positivo para las oportunidades y definiciones de impacto negativo para las amenazas. Se pueden utilizar para la probabilidad y el impacto términos descriptivos (como muy alto, alto, medio, bajo y muy bajo) o valores numéricos. Cuando se utilizan valores numéricos, estos pueden ser multiplicados para dar una puntuación de probabilidad de impacto para cada riesgo, lo que permite que la prioridad relativa de los riesgos individuales sea evaluada dentro de cada nivel de prioridad.

Gráfico 5 Matriz de probabilidad e impacto con esquema de puntuación. Autor: (PMBOK_V6, 2017)

Probabilidad	Amenazas					Oportunidades					Probabilidad
	Muy Bajo 0,05	Bajo 0,09	Moderado 0,18	Alto 0,36	Muy Alto 0,72	Muy Bajo 0,72	Bajo 0,36	Moderado 0,18	Alto 0,09	Muy Alto 0,05	
Muy alta 0,90	0,05	0,09	0,18	0,36	0,72	0,72	0,36	0,18	0,09	0,05	Muy alta 0,90
Alta 0,70	0,04	0,07	0,14	0,28	0,56	0,56	0,28	0,14	0,07	0,04	Alta 0,70
Mediana 0,50	0,03	0,05	0,1	0,2	0,4	0,4	0,2	0,1	0,05	0,03	Mediana 0,50
Baja 0,30	0,02	0,03	0,06	0,12	0,24	0,24	0,12	0,06	0,03	0,02	Baja 0,30
Muy Baja 0,10	0,01	0,01	0,02	0,04	0,08	0,08	0,04	0,02	0,01	0,01	Muy Baja 0,10
Impacto negativo					Impacto positivo						
Muy Bajo 0,05	Bajo 0,10	Moderado 0,20	Alto 0,40	Muy Alto 0,80	Muy Alto 0,80	Alto 0,40	Moderado 0,20	Bajo 0,10	Muy Bajo 0,05		

2.1.7 Análisis Cualitativo de riesgos

Implica analizar los riesgos y evaluar la probabilidad de que ocurran, y que impacto habría si ocurrieran. En función de eso priorizas los riesgos. Lo primero que haces al analizar los riesgos es tomar la lista de riesgos identificada y determinar cuáles son los riesgos más importantes, ya que sólo a éstos se les realizará un análisis más profundo y se le planificarán respuestas (Buchtik, 2015).

La evaluación del riesgo individual mediante el análisis de riesgo cualitativo evalúa la probabilidad de que ocurra cada riesgo y el efecto de cada riesgo individual en los objetivos del proyecto. como tal, no aborda directamente el riesgo general para los objetivos del proyecto que resulta del efecto combinado de todos los riesgos y sus interacciones potenciales entre sí. Sin embargo, esto puede lograrse mediante el uso de técnicas de análisis de riesgos cuantitativas (Practice Standard for Project Risk Management, 2009).

Se debe concentrar en los riesgos prioritarios dado que tiene que haber un equilibrio entre el costo y el beneficio de realizar el análisis. Llevar este análisis con el equipo no es complicado, pero lleva tiempo y esfuerzo, tiene un costo. Con esto no quiere decir que no se lo haga, todo lo contrario, este es un proceso obligatorio, no relacionarlo amenazaría el éxito del proyecto, no se debe subestimar el análisis ni tampoco analizar más allá de lo necesario. Analiza los riesgos siendo consistente con la importancia y la criticidad del proyecto (Buchtik, 2015).

2.1.8 Análisis Cuantitativo de Riesgos

Realizar el Análisis Cuantitativo de Riesgos no es necesario para todos los proyectos. La realización de un análisis profundo depende de la disponibilidad de datos de alta calidad sobre los riesgos individuales del proyecto y otras fuentes de incertidumbre, así como de una sólida línea base del proyecto subyacente para el alcance, el cronograma y el costo. El análisis cuantitativo de riesgos por lo general requiere de un software de riesgos especializado y pericia en el desarrollo y la interpretación de los modelos de riesgo. Además, consume tiempo y costos adicionales. Es probablemente apropiado para proyectos grandes o complejos, proyectos estratégicamente importantes, proyectos para los cuales es un requisito contractual o proyectos en los que un interesado clave lo requiere. El análisis cuantitativo de riesgos es el único método confiable para evaluar el riesgo general del proyecto a través de la evaluación del efecto global sobre los resultados del proyecto de todos los riesgos individuales del proyecto y otras fuentes de incertidumbre (PMBOK_V6, 2017).

Realizar el Análisis Cuantitativo de Riesgos utiliza la información sobre los riesgos individuales del proyecto que han sido evaluados al momento de realizar el Análisis Cualitativo de Riesgos como que presentan un potencial significativo para afectar los objetivos del proyecto (PMBOK_V6, 2017).

Es indagar, mediante algún modelo matemático, el efecto de los riesgos y sus interacciones, sobre los objetivos del costo y del cronograma del proyecto. En este paso analizamos más objetivamente o numéricamente. El análisis numérico parte de la lista de riesgos que requieren un análisis mayor (Buchtik, 2015).

Este análisis brinda un enfoque adicional para tomar decisiones sobre los riesgos. Asume que el riesgo opera de modo previsible, que los factores que llevan a los eventos de riesgo se pueden identificar, categorizar, controlar, y que con más análisis tendrás más respuestas. Cuidado porque dicha hipótesis no aporta si hay cisnes negros. Sin embargo, recuerda siempre que los proyectos operan con personas y el comportamiento humano a veces no se puede predecir y no siempre es intuitivo (Nassim, 2010).

Cuantificar un riesgo es determinar los valores posibles que puede tomar una variable de riesgo (posibles resultados), así como la probabilidad de que ocurra cada uno de esos

valores. Una variable puede ser el costo de una tarea, la rentabilidad, el número de ventas, la duración de una actividad, entre otros (Buchtik, 2015).

Gráfico 6 Tipos de análisis de riesgos. Autor: (Buchtik, 2015)

TIPOS de ANÁLISIS de RIESGOS		
ANÁLISIS CUALITATIVO		ANÁLISIS NUMÉRICO
¿Qué implica?	Priorizar los riesgos individuales evaluando la probabilidad de que ocurran, y el impacto si ocurren.	Analizar numéricamente el efecto de los riesgos sobre el proyecto general, como un todo. Asigna una calificación numérica a cada riesgo o a un acumulado de ellos.
¿Cuál es su precisión?	Es subjetivo. Se basa en impresiones.	Es más objetivo. Se basa en datos, modelos y/o simulación.
¿Es obligatorio?	Sí.	No. No siempre es necesario
¿Cuándo se precisa?	Siempre, en todo proyecto.	Cuando se requiere un análisis de mayor certeza, o numérico.
¿Cuál es su ventaja?	Es relativamente fácil, rápido y barato. Es suficiente para la mayoría de los proyectos.	Es más exacto. Sirve para saber qué probabilidad hay de terminar en una fecha dada o a un costo determinado. Sirve para ver que riesgos impactan más en objetivos del proyecto.
¿Cuál es su desventaja?	No es tan exacto como el análisis numérico.	Lleva más tiempo. Requiere crear modelos, usar software, saber más, tener datos históricos relativamente recientes y de calidad. Es más caro y complejo.
¿Qué debes saber?	Probabilidad básica.	Probabilidad y estadística avanzada, tipos de distribuciones, modelado, simulación y programación del tiempo.
¿Qué haces primero?	Este se hace primero.	Esto se hace después, si se hace.
¿Qué tipos de proyecto lo usan?	Todo tipo de proyectos.	En general los proyectos más grandes, complejos, de alto costo, y megaproyectos.

2.2 Línea Base

2.2.1 Línea Base del Alcance

La línea base del alcance es la versión aprobada de un enunciado del alcance (PMBOK_V6, 2017).

2.2.2 Línea Base del Cronograma

Una línea base del cronograma consiste en la versión aprobada de un modelo de programación que sólo puede cambiarse mediante procedimiento formales de control de cambios y que se utiliza como base de comparación con los resultados reales. Es aceptada y aprobada por los interesados adecuados como la línea base del cronograma, con fechas de inicio de la línea base y fechas de finalización de la línea base. Durante el monitoreo y control, las fechas aprobadas de la línea base se comparan con las fechas reales de inicio y finalización para determinar si se han producido desviaciones (PMBOK_V6, 2017).

2.2.3 Línea Base del Costo.

La línea base de costos es la versión aprobada del presupuesto del proyecto con fases de tiempo, excluida cualquier reserva de gestión, la cual sólo puede cambiarse a través de procedimientos formales de control de cambios. Se utiliza como base de comparación con los resultados reales. La línea base de costos se desarrolla como la suma de los presupuestos aprobados para las diferentes actividades del cronograma (PMBOK_V6, 2017).

2.3 Estimaciones

2.3.1 Estimación Análoga

La estimación análoga es una técnica para estimar la duración o el costo de una actividad o de un producto utilizando datos históricos de una actividad o proyecto similar. La estimación análoga utiliza parámetros de un proyecto anterior similar, tales como duración, presupuesto, tamaño, peso y complejidad, como base para estimar los mismos parámetros o medidas para un proyecto futuro. Cuando se trata de estimar duraciones, esta técnica utiliza la duración real de proyectos similares anteriores como base para estimar la duración del proyecto actual. Es un método de estimación del valor bruto, que en ocasiones se ajusta en función de las diferencias conocidas en cuanto a la complejidad del proyecto (Practice Standard for Project Estimating, 2011).

La estimación análoga de la duración se emplea a menudo para estimar la duración de un proyecto cuando se dispone de escasa información de detalle sobre el mismo (Practice Standard for Project Estimating, 2011).

Por lo general, la estimación análoga es menos costosa y requiere menos tiempo que otras técnicas, pero también es menos exacta. La estimación análoga de duraciones se puede aplicar a un proyecto en su totalidad o a partes del mismo, y puede utilizarse en combinación con otros métodos de estimación. La estimación análoga es más fiable cuando las actividades anteriores son hechos similares, no sólo en apariencia, y cuando los miembros del equipo del proyecto responsables de efectuar las estimaciones poseen la apariencia necesaria (Practice Standard for Project Estimating, 2011).

2.3.2 Estimación Paramétrica

La estimación paramétrica es una técnica de estimación en la que se utiliza un algoritmo para calcular el costo o la duración con base a datos históricos y otras variables (p.ej., metros cuadrados de construcción) para calcular una estimación de los parámetros de una actividad tales como costo, presupuesto y duración (Practice Standard for Project Estimating, 2011).

Las duraciones pueden determinarse cuantitativamente multiplicando la cantidad de trabajo a realizar por el número de horas por unidad de trabajo. Por ejemplo, la duración de un proyecto de diseño puede estimarse multiplicando el número de planos por la cantidad de horas laborales necesarias para cada plano; o para una instalación de cable, multiplicando los metros de cable por la cantidad de horas laborales por metro. Si el recurso asignado es capaz de instalar 25 metros de cable por hora, la duración requerida para instalar 1.000 metros sería de 40 horas (1.000 metros divididos entre 25 metros por hora) (Practice Standard for Project Estimating, 2011).

Con esta técnica se pueden lograr niveles superiores de exactitud, en función de la sofisticación y de los datos subyacentes que utilice el modelo. La estimación paramétrica del cronograma se puede aplicar a un proyecto en su totalidad o a partes del mismo, y se puede utilizar en conjunto con otros métodos de estimación (Practice Standard for Project Estimating, 2011).

2.3.3 Estimación Basada En Tres Valores o PERT

La exactitud de las estimaciones de la duración o el costo por único valor puede mejorarse si se tienen en cuenta la incertidumbre y el riesgo. El uso de estimaciones basadas en tres valores ayuda a definir un rango aproximado de duración de una actividad:

- **Más probable (tM).** Esta estimación se basa en la duración o el costo de la actividad, en función de los recursos que probablemente le sean asignados, de su productividad, de las expectativas realistas de disponibilidad para la actividad, de las dependencias de otros participantes y de las interrupciones.
- **Optimista (tO).** Estima la duración o el costo de la actividad sobre la base del análisis del mejor escenario para esa actividad.
- **Pesimista (tP).** Estima la duración o el costo sobre la base del análisis del peor escenario para esa actividad.

Se puede calcular la duración o el costo esperado, tE y/o cE, en función de la distribución asumida de los valores dentro del rango las tres estimaciones. Dos de las fórmulas más utilizadas son las distribuciones triangular y beta. Las fórmulas son las siguientes:

Distribución triangular. $tE = (tO + tM + tP) / 3$

$$cE = (cO + cM + cP) / 3$$

Distribución beta. $tE = (tO + 4tM + tP)/6$

$$cE = (cO + 4cM + cP)/6$$

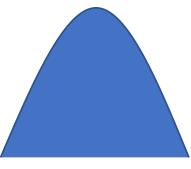
La distribución triangular y distribución beta se utiliza cuando existen datos históricos insuficientes o cuando se usan datos subjetivos. Las estimaciones de duración o costo basada en tres valores con una distribución determinada proporcionan una duración o costo esperado y despejan el grado de incertidumbre sobre la duración o el costo esperado (Practice Standard for Project Estimating, 2011).

2.4 Distribuciones Probabilísticas

Lo primero a considerar al desarrollar un modelo que cuantifique los riesgos es que dejas de hablar de estimaciones determinísticas para hablar de estimaciones probabilísticas. Ya no estimas sólo un valor sino un rango. Ese rango estará contenido entre un valor mínimo, un máximo y a veces uno más probable, contenido entre los anteriores. Entre sus extremos optimistas y pesimistas la distribución tiene un 100% de probabilidad. Una vez que cuantificas un riesgo este se puede describir mediante una distribución probabilística, la cual representa la incertidumbre de una variable, establece el rango que pueden tomar los valores y la probabilidad de que ocurra cada valor del rango.

En una distribución, el eje “X” representa los valores de duración de las actividades (tiempo) a simular, o sus valores del costo. El eje “Y” representa la probabilidad de que ocurra cada valor (o la cantidad de veces que se produjo ese valor al simular). La forma de la distribución se muestra cómo se comporta la variable. Hay varias distribuciones que se usan en la gestión de riesgos. Hay dos tipos de distribuciones, las continuas y las discretas. Las discretas pueden tomar sólo ciertos valores finitos, las continuas pueden tomar cualquier valor del rango, y representar la incertidumbre de los valores de las duraciones de las actividades o de su costo (Buchtic, 2023).

Gráfico 7 Distribuciones probabilísticas usadas en el análisis numérico de riesgos. Autor: (Buchtic, 2023)

DISTRIBUCIONES COMUNES			
Distribución	Representación	¿Cuándo se usa?	Ejemplo de uso
UNIFORME		Cuando tienes la mayor incertidumbre. Todos los valores del rango son igualmente probables, incluso los valores extremos.	Un equipo donde no se conocen entre ellos. Con una actividad con la cual no tienes experiencia. Ésta no es muy útil.
TRIANGULAR		Tienes un poco más de información. El valor del medio también influye en la ponderación	Hay una estimación pero no es tan buena como para confiar más en un valor dado. Por eso se promedian los tres valores del rango.
NORMAL O GAUSEANA		Hay estimaciones con mayor confianza; puedes determinar una media, y confiar que los valores se agruparán en torno a ella.	Tienes más conocimiento sobre la actividad. La organización posee información estadística sobre rangos de plazo o costo de la actividad o similares.

2.4.1 Distribuciones Probabilísticas Para Cuantificar los Riesgos.

El análisis numérico de riesgos examina la incertidumbre que hay sobre las variables de tiempo o de costo del plan del proyecto. La incertidumbre se modela a través de distribuciones de probabilidad. Según como se comporte el fenómeno que se quiera modelar, es el tipo de distribución a elegir.

Un modelo es una presentación de la realidad que vas a analizar para intentar entender mejor qué sucederá en el futuro. Sobre ese modelo ejecutarás una simulación. El modelo tiene valores de entrada y de salida. Los valores de salida se visualizan mediante las distribuciones de probabilidad, las cuales permiten definir distintos tipos de incertidumbre (Buchtic, 2023).

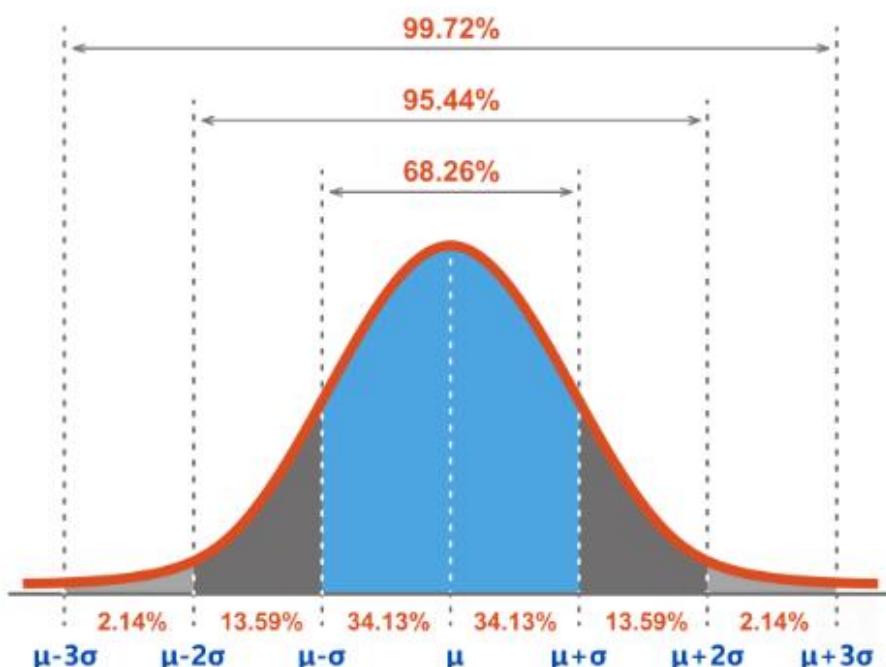
2.4.2 Distribución Normal

Usa la media y la desviación estándar como parámetros de la distribución. La media se representa con el símbolo μ , y la desviación estándar con el σ .

La media (μ) define el valor alrededor del cual se centra la curva que es simétrica.

La desviación estándar (σ) indica que la dispersión tiene los valores en torno a la media. En general, a mayor dispersión, mayor riesgo (Buchtic, 2023).

Gráfico 8 Distribuciones probabilísticas normal Autor: (Buchtic, 2023)



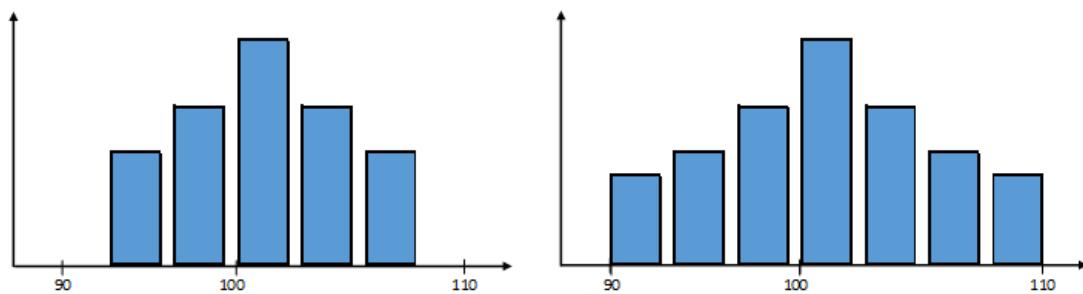
Esta distribución representa que la mayoría de los resultados se dan cerca del promedio (μ). Es simétrica respecto del valor medio. La mitad del área bajo la curva está a la derecha del punto máximo -- la media, y la otra mitad del área está a su izquierda. En el Gráfico 8 se observa que: (Buchtic, 2023).

- El 68.26% de los valores se acerca a la media se calcula: $34.13\% + 34.13\%$
- El 95.44% se da entre $+2\sigma$ y -2σ $(34.13\% + 13.6\%) * 2$
- El 99.72 % se da entre $+3\sigma$ y -3σ $(34.13\% + 13.6\% + 2.14) * 2$

En la vida real, la distribución normal se podría usar cuando por ejemplo en un departamento, todos sus empleados ganan más o menos lo mismo, todos están cerca del promedio; o en un proyecto, los materiales de las columnas de los diferentes proveedores todos tienen un precio similar y están cerca de la media.

Los adversos al riesgo tienden a preferir una distribución angosta o de menor riesgo donde sus valores se concentran más en un lugar. Los arriesgados prefieren una distribución más amplia o de mayor riesgo. El Gráfico 9 muestra a la izquierda una distribución de menor riesgo y a la derecha una de mayor riesgo (Buchtic, 2023).

Gráfico 9 Ej. Distribuciones normales con distinto riesgo Autor: (Buchtic, 2023)

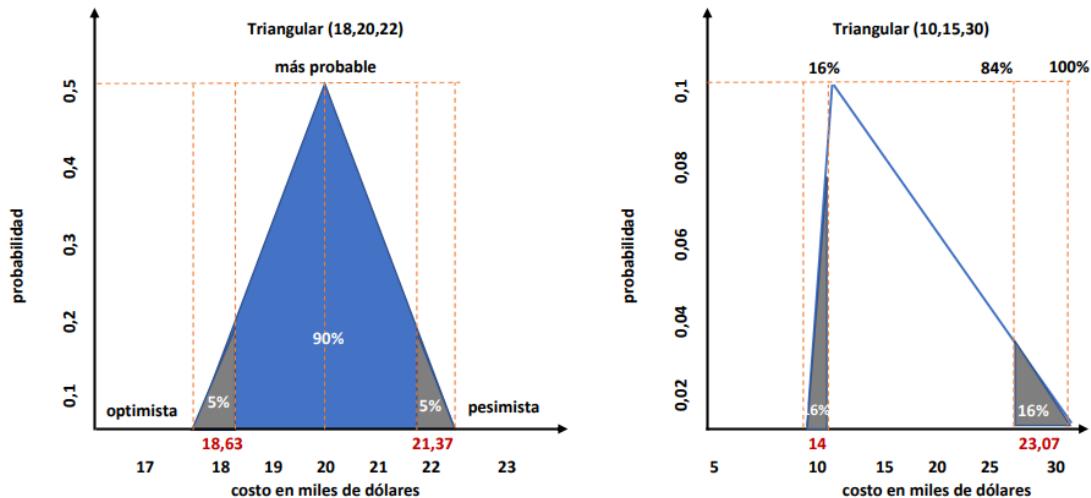


2.4.3Distribución Triangular

Si la curva indica un valor posible mínimo de \$10.000 en el costo de una tarea, el valor más probable de \$15.000 y el máximo de \$30.000 (Gráfico 10 a la derecha), ¿Cuál de los tres valores usarás para estimar el costo de la tarea? Depende de la actitud frente al

riesgo del decidor. Un arriesgado se inclinará por un valor cercano mínimo o \$10.000 porque es optimista, Un adverso al riesgo elegirá un valor próximo a los \$30.000 ya que intentará reducir la probabilidad de no cumplir con lo estimado y tratará de cubrirse al máximo. Un decidor neutro estimará \$15.000 (Buchtic, 2023).

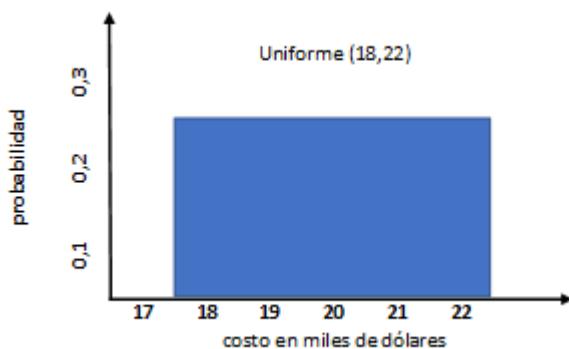
Gráfico 10 Distribuciones triangulares. Autor: (Buchtic, 2023)



2.4.4Distribución Uniforme

En esta distribución los valores tienen toda la misma probabilidad de ocurrir, es uniforme. Se usa si hay mucha incertidumbre, si no se puede optar por un valor más probable y sólo se conocen los extremos (peor y mejor caso) en los cuales se puede mover la variable (Buchtic, 2023).

Gráfico 11 Distribuciones uniforme. Autor: (Buchtic, 2023)

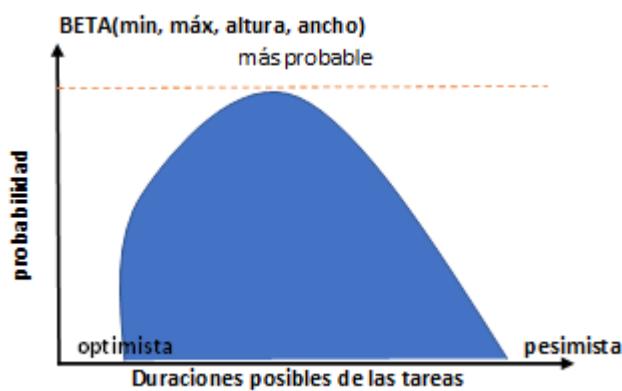


2.4.5 Distribución Beta o Pert

Puede ser simétrica o asimétrica y le da más peso al valor más probable. La mayoría de los números están cerca del valor más probable o del valor esperado.

También se llama PERT. Tiene 4 parámetros: mínimo, máximo, y dos valores sobre la forma de la distribución que dice que tan alta y ancha es la misma (Buchtic, 2023).

Gráfico 12 Distribución beta Autor: (Buchtic, 2023)



2.5 Lógica Difusa

Es una lógica multivaluada que permite representar matemáticamente la incertidumbre y la vaguedad, proporcionando herramientas formales para su tratamiento. “Cuando aumenta la complejidad, los enunciados precisos pierden su significado y los enunciados útiles pierden precisión” (Lógica Difusa, 2017).

Básicamente, cualquier problema del mundo real puede resolverse como dado un conjunto de variables de entrada, obtener un valor adecuado de variables de salida. La lógica difusa permite establecer este mapeo de una forma adecuada, atendiendo a criterios de significado y no de precisión (Lógica Difusa, 2017).

2.5.1 Lógica Difusa en Inteligencia Artificial.

La lógica difusa es una rama de la inteligencia artificial que le permite a una computadora analizar información del mundo real en una escala entre lo falso y lo verdadero, manipula conceptos vagos, como “caliente” o “frío”. Los sistemas basados en lógica difusa imitan

la forma en que toman decisiones los humanos, con la ventaja de ser mucho más rápidos. Consiste en la aplicación de la lógica difusa con la intención de imitar el razonamiento humano en la programación de computadoras. Con la lógica convencional, las computadoras pueden manipular valores estrictamente duales, como verdadero/falso, sí/no. En la lógica difusa, se usan modelos matemáticos para representar nociones subjetivas, como: optimista/más probable/pesimista, frío/caliente/tibio, medio/bajo/alto para valores concretos que puedan ser manipuladas por los ordenadores. En este paradigma, también tiene un especial valor en la gestión de riesgos cuantitativos tanto en la distribución Beta – Pert así como en la Estimación Basada En Tres Valores o PERT (Lógica Difusa, 2017).

2.6 Machine Learning

Machine Learning permite elaborar un modelo de datos, definido por ciertos datos de entrada y a través de la programación optimizar los parámetros del modelo y usarlos para predecir salidas de datos futuros. Machine Learning utiliza la teoría estadística para la elaboración de modelos, esto es porque se está realizando una inferencia a través de una muestra; y además usa la programación computacional para optimizar los parámetros del modelo mediante algoritmos eficientes tanto para la optimización del modelo como para su uso en la predicción (Alpaydin, 2014).

2.6.1 Aprendizaje Supervisado

En Machine Learning tenemos diferentes métodos de ajuste de los parámetros de un modelo, lo que se conoce como aprendizaje. En el aprendizaje supervisado, para cada ejemplo tenemos datos de entrada y los datos de salida deseados. Además, asumimos que existe un “supervisor” que puede proveer de los datos de salida deseados para un determinado dato de entrada. El aprendizaje se refiere al ajuste de parámetros de tal manera que el modelo pueda realizar las predicciones más precisas de los datos, en otras palabras, el aprendizaje se refiere al mejoramiento de acuerdo con un criterio de desempeño, que puede ser la predicción, clasificación entre otras palabras. Un modelo de aprendizaje supervisado busca aprender de un conjunto de datos de entrada y predecir futuros datos de salida de manera precisa (Flach P, 2012).

2.6.2 Aprendizaje No Supervisado

En este tipo de aprendizaje no existe un supervisor, es decir no se cuentan con datos de entrada y salida apareados, solo existen datos de entrada. Para aprender se trata de encontrar regularidades en los datos de entrada. En estos casos se busca una estructura dentro de los datos de entrada, en la que se puedan encontrar patrones en los datos que son más comunes, con esto se puede identificar qué es lo que ocurre con más frecuencia, desde el punto de vista estadístico este proceso es la estimación de la densidad. Existen varios métodos de estimación de la densidad como “clustering”, y asociación (Alpaydin, 2014).

2.6.3 Modelo

Un modelo es una representación de la realidad y te ayuda a cuantificar el riesgo para luego decidir si vale la pena tomarlo o no, o qué acción conviene aplicar. Si hay un 15% de probabilidad de terminar el proyecto tarde, costando \$1,500 más de lo presupuestado, quizás estés dispuesto a asumir el riesgo. Pero si la fecha de fin fijada tiene un 40% de posibilidad de no cumplirse y una multa de \$150.000 por incumplimiento, seguramente no estarás dispuesto a asumirlo, o estarás más dispuesto a mitigarlo asignando más recursos que bajen la probabilidad de retraso. Mediante fórmulas, funciones y datos, un modelo representa la relación que hay entre las variables de entrada y de salida (Buchtik, 2015).

La Variación muestra cómo varían los datos. Por ejemplo, cómo varió en dos años a través de diferentes proyectos la duración de tareas similares, los costos de los materiales de una tarea, entre otros. La variación se representa mediante distribuciones de probabilidad que reflejan el comportamiento esperado de las variables de entrada. Una distribución indica la frecuencia de los posibles valores que puede tomar la variable de entrada. Por ejemplo, si hay datos históricos que muestran cuando duró una tarea particular cada vez que se realizó a lo largo de dos años, y se sabe que esos datos tienen un patrón de distribución triangular, entonces se usará una distribución triangular en el modelo (Buchtik, 2015).

Con las variables de entrada se ejecuta el modelo. Indicas cuántas veces quieres simular el modelo, es decir la cantidad de escenarios o iteraciones, y se generan las variables de

salida o resultados. El software de simulación muestra cuánto dura cada tarea simulada en cada iteración. Por ejemplo, la tarea 25 demoró 10 días la primera vez que se ejecutó, 12 días la segunda vez, 15 días la tercera vez, y así sucesivamente si el modelo iteró 1,000 veces, se puede saber la probabilidad de duración de la tarea, ya que hay información sobre cuantas veces demoró 10 días, cuántas veces demoró 15 días, etc., a lo largo de mil escenarios (Buchtik, 2015).

La simulación da como resultado la probabilidad de terminar el proyecto en una fecha dada, y el riesgo asociado a cada fecha de fin posible que surge al simular. Lo mismo para el costo. También indica cuántas veces una tarea estuvo en el camino crítico, entre otros. Determina el riesgo general del proyecto (Buchtik, 2015).

2.7 Algoritmos de machine learning relevantes que se aplican en la gestión de riesgos cuantitativos.

2.7.1 Estudio McKinsey, “Notes from the AI frontier insights from hundreds of use cases.”

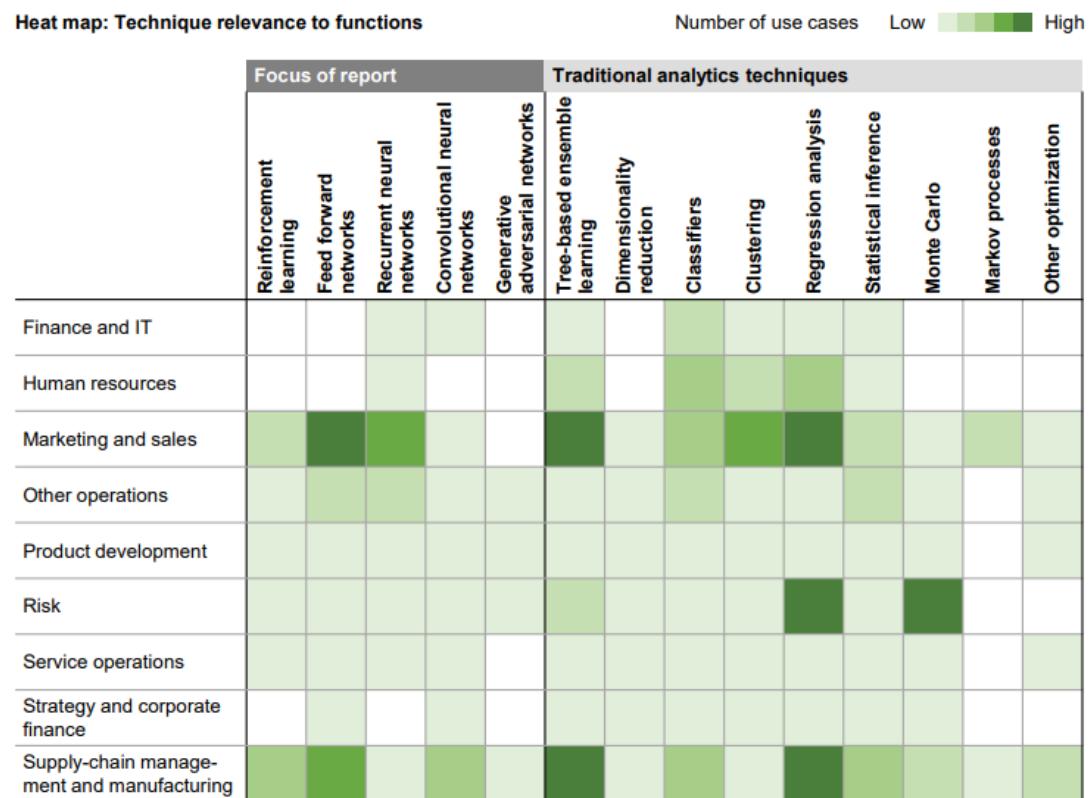
Un análisis de más de 400 casos de uso en 19 industrias y nueve funciones comerciales destaca el amplio uso y el importante potencial económico de las técnicas avanzadas de Inteligencia Artificial IA (McKinsey&Company, 2018).

La inteligencia artificial (IA) se destaca como una tecnología transformadora de nuestra era digital, y su aplicación práctica en toda la economía está creciendo a buen ritmo. Para este informe mapeamos tanto el análisis tradicional como las nuevas técnicas de “aprendizaje profundo” y los problemas que pueden resolver en más de 400 casos de uso específicos en empresas y organizaciones. Basándonos en la investigación del McKinsey Global Institute y la experiencia aplicada con IA de McKinsey Analytics, evaluamos tanto las aplicaciones prácticas como el potencial económico de las técnicas avanzadas de IA en todas las industrias y funciones comerciales. Nuestros hallazgos resaltan el potencial sustancial de aplicar técnicas de Deep learning para casos de uso en toda la economía, pero también vemos algunas limitaciones y obstáculos continuos, junto con oportunidades futuras a medida que las tecnologías continúan su avance. En última instancia, el valor de IA no se encuentra en los modelos en sí, sino en la capacidad de las empresas para aprovecharlos (McKinsey&Company, 2018).

Es importante resaltar que, aunque vemos potencial económico en el uso de técnicas de IA, el uso de datos siempre debe tener en cuenta preocupaciones que incluyen la seguridad de los datos, la privacidad y posibles problemas de sesgo (McKinsey&Company, 2018).

El estudio realizado por McKinsey en una parte muestra un mapa de calor en el cual consta la aplicabilidad de las diferentes técnicas variaba según los sectores y funciones comerciales.

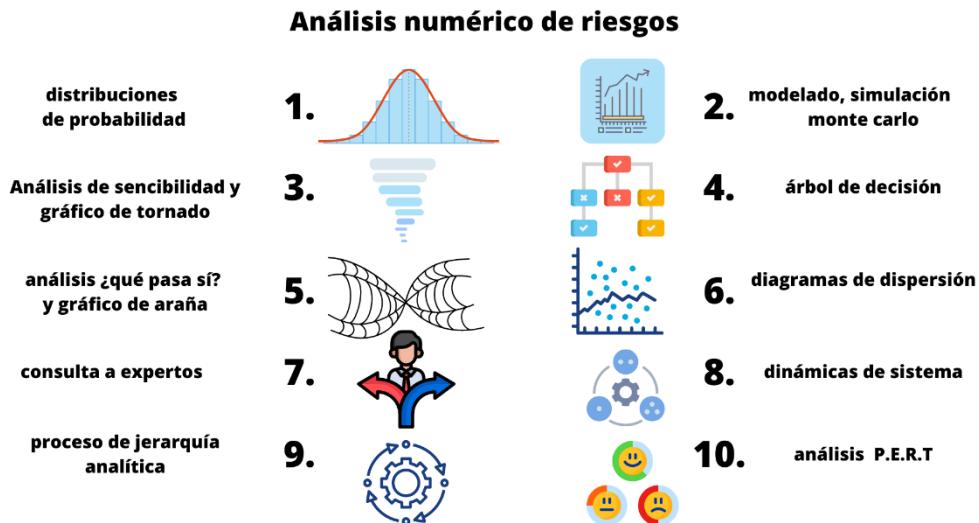
Gráfico 13 Mapa de calor: Relevancia de la técnica para las funciones Autor: (McKinsey&Company, 2018)



2.7.2 Herramientas para análisis numérico de riesgos

Para analizar el riesgo del costo del proyecto usas estimaciones de costo como variables de entrada. Para analizar el riesgo del cronograma usas estimaciones de la duración de las actividades como variables de entrada.

Gráfico 14 Caja de herramientas para analizar los riesgos numéricamente Autor: (Buchtik L. , 2015)



2.7.3 Algoritmos de machine learning seleccionados para aplicar en la gestión de riesgos cuantitativos.

La investigación realizada por McKinsey&Company en el año 2018, así como Liliana Buchtik en su libro Secretos Para Dominar La Gestión de Riesgos En Proyectos publicado en el año 2015. Ambos nos permiten ver que para el sector de Riesgos (Risk) aplican los siguientes algoritmos de machine learning:

- a) Monte Carlo
- b) Árboles de decisión
- c) Análisis de regresión

2.8 Monte Carlo

2.8.1 ¿Qué es la simulación Monte Carlo?

La simulación Monte Carlo, también conocida como el método de Monte Carlo o simulación de probabilidad múltiple, es una técnica matemática que se utiliza para estimar los posibles resultados de un evento incierto. El método Monte Carlo fue inventado por John Von Neumann y Stanislaw Ulam durante la Segunda Guerra Mundial para mejorar la toma de decisiones en

condiciones inciertas. Lleva el nombre de una conocida ciudad de casinos, llamada Mónaco, ya que el elemento de azar es fundamental para el enfoque de modelado, similar a un juego de ruleta.

Desde su introducción, las simulaciones Monte Carlo han evaluado el impacto del riesgo en muchos escenarios de la vida real, como la inteligencia artificial, los precios de acciones, la previsión de ventas, la gestión de proyectos y la fijación de precios. También proporcionan una serie de ventajas sobre modelos predictivos con entradas fijas, como la capacidad de realizar análisis de sensibilidad o calcular la correlación de entradas. El análisis de sensibilidad permite a los tomadores de decisiones ver el impacto de las entradas individuales en un resultado específico y la correlación les permite comprender las relaciones entre cualquier variable de entrada (IBM, 2022).

2.8.2 ¿Cómo funciona la simulación Monte Carlo?

A diferencia de un modelo de pronóstico normal, la simulación Monte Carlo predice un conjunto de resultados basados en un rango estimado de valores frente a un conjunto de valores de entradas fijos. En otras palabras, una simulación Monte Carlo crea un modelo de posibles resultados al utilizar una distribución de probabilidades, como una distribución uniforme o normal, así como triangular y/o beta Pert, para cualquier variable que tenga una incertidumbre inherente. Luego, recalcula los resultados una y otra vez, cada vez mediante un conjunto diferente de números aleatorios entre los valores mínimo y máximo. En un experimento típico de Monte Carlo, este ejercicio se puede repetir miles de veces para producir una gran cantidad de resultados probables.

Las simulaciones Monte Carlo también se utilizan para predicciones a largo plazo debido a su precisión. A medida que aumenta el número de entradas, también crece el número de pronósticos, lo que le permite proyectar los resultados más lejos en el tiempo con más precisión. Cuando se completa una simulación Monte Carlo, se obtiene un rango de resultados posibles con la probabilidad de que ocurra cada resultado.

Un ejemplo simple de una simulación Monte Carlo es considerar el cálculo de la probabilidad de lanzar dos dados estándar. Hay 36 combinaciones de tiradas de dados. En base a esto, puede calcular manualmente la probabilidad de un resultado en particular. Con una simulación Monte Carlo, puede simular lanzar los dados 10,000 veces (o más) para lograr predicciones más precisas (IBM, 2022).

2.8.3 Cómo utilizar los métodos de Monte Carlo

Independientemente de la herramienta que utilice, las técnicas de Monte Carlo implican tres pasos básicos:

1. Configure el modelo predictivo, identificando tanto la variable dependiente que se va a predecir como las variables independientes (también conocidas como variables de entrada, riesgo o predictores) que impulsarán la predicción.
2. Especifique las distribuciones de probabilidades de las variables independientes. Utilice datos históricos y/o el juicio subjetivo del analista para definir un rango de valores probables y asigne ponderaciones de probabilidad para cada uno.
3. Ejecute simulaciones repetidamente, generando valores aleatorios de las variables independientes. Haga esto hasta que se obtenga suficientes resultados para formar una muestra representativa del número casi infinito de combinaciones posibles.

Para ejecutar tantas simulaciones Monte Carlo como desee modificando los parámetros subyacentes que utiliza para simular los datos. Sin embargo, también querrá calcular el rango de variación dentro de una muestra al calcular la varianza y la desviación estándar que son medidas de dispersión de uso común (IBM, 2022).

La técnica Monte Carlo es un método de muestreo que genera valores de entrada al azar que se usan durante una simulación. Se considera una mejor práctica internacional. Para cada tarea que simule, en vez de usar un sólo valor como el más probable, por ejemplo, usa varios valores posibles al azar de su distribución; así combina y obtiene todos los resultados posibles del proyecto o fase. Por ejemplo, si se sabe que el costo de la tarea es A estará entre \$ 50.000 y \$ 100.000, y el costo de la tarea B entre \$90.000 y \$100.000, la simulación tomará, para cada tarea, valores al azar entre los rangos indicados. Luego de simular, estos valores deben respetar la distribución de probabilidad configurada ya que usará los valores en función de su probabilidad de ocurrencia, los más probables tendrán más peso (Buchtik, 2015).

La simulación permite evaluar el efecto de los riesgos en el proyecto para predecir cómo se puede comportar éste, y ver qué tan realista es el presupuesto y/o el cronograma para poder ajustarlo antes de comprometerse (Buchtik, 2015).

Ventajas:

- Es una potente herramienta que permite analizar muchos escenarios para observar sus efectos.

- Permite pronosticar fechas y costos más realistas, y ver si deberías modificar el costo o la fecha.
- Ayuda a determinar la contingencia.
- Muestra la probabilidad de que ciertas actividades estén en el camino crítico.

Desventajas:

- Aplica sólo a costo y tiempo
- Requiere tener suficientes datos históricos y el uso de software
- Demanda tiempo para verificar que el Gantt siga las buenas prácticas de correr la simulación (ej. sin hitos, sin restricciones, sin dependencias, sin contingencias en las tareas, minimizando adelantos o retrasos, etc.) y con los recursos y costos cargados.
- Requiere saber probabilidad, estadística.
- Asume que el desempeño futuro será similar al desempeño pasado y eso no siempre es así.

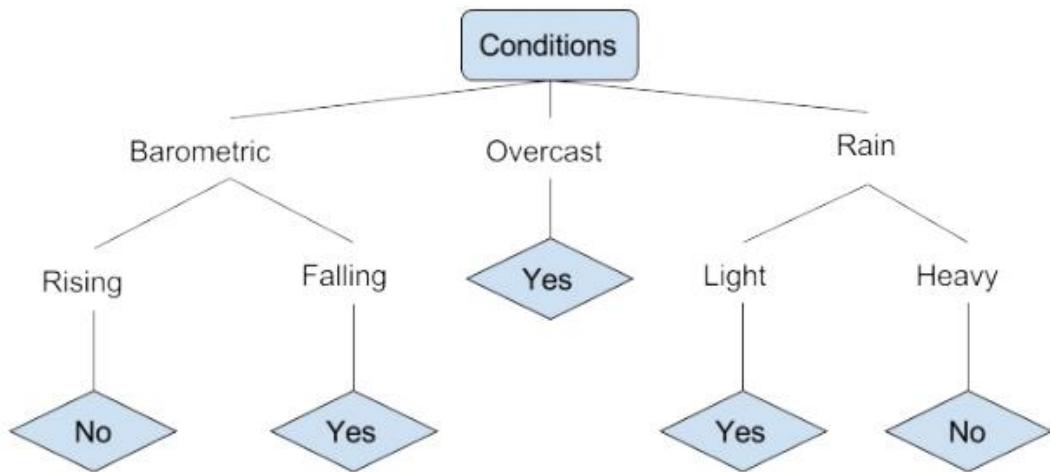
2.9 Árboles de decisión

Los árboles de decisión se emplean para representar visualmente decisiones y mostrar o informar la toma de decisiones. Al trabajar con machine learning y data mining, los árboles de decisión se utilizan como un modelo predictivo. Estos modelos asignan observaciones sobre datos a concluirse sobre el valor objetivo de los datos. El objetivo del aprendizaje del árbol de decisiones es crear un modelo que prediga el valor de un objetivo en función de las variables de entrada (Brian Boucheron & Lisa Tagliaferri, 2022).

En el modelo predictivo, los atributos de los datos que se determinan a través de las observaciones están representados por las ramas, mientras que las conclusiones sobre el valor objetivo de los datos se presentan en las hojas (Brian Boucheron & Lisa Tagliaferri, 2022).

Veamos un ejemplo de varias condiciones que pueden determinar si alguien debe o no ir a pescar. Esto incluye el clima condiciones, así como las condiciones de presión barométrica (Brian Boucheron & Lisa Tagliaferri, 2022).

Gráfico 15 Ejemplo de árbol de decisiones de pesca Autor: (Brian Boucheron & Lisa Tagliaferri, 2022)



En el árbol de decisión simplificado anterior, el ejemplo se ha clasificado a través del árbol hasta el nodo de la hoja apropiado. Esto luego devuelve la clasificación asociada con la hoja en particular, que en este caso es un Sí o un No. El árbol clasifica las condiciones de un día en función de si es adecuado o no ir a pescar (Brian Boucheron & Lisa Tagliaferri, 2022).

Ventajas:

- Es fácil de entender y visual.
- Hay software para dibujarlo lo cual presenta profesionalmente las alternativas y resultados, y hace los cálculos automáticamente
- Permite seleccionar la opción con la mejor ganancia o con el menor costo.

Desventajas:

- Si no hay fundamentos de las estimaciones de probabilidad de ocurrencia y del costo y ganancia de cada alternativa, el resultado no es realista.
- Es impráctico si hay muchos eventos de riesgo porque la cantidad total de resultados posibles aumenta espontáneamente.

2.10 Análisis de regresión

El análisis de regresión es una técnica usada para modelar la relación entre variables. Se desea establecer como una o varias variables dependientes se comportan respecto a una o más variables independientes. Mediante esta técnica podemos obtener información sobre como una variable de interés Y, variable dependiente, varia cuando una de las independientes lo hace (Carlos N. Bouza, 2018).

Desde el punto de vista formal el análisis de regresión se dedica a estimar la esperanza condicional de Y para dadas las independientes. Esto es como varía el promedio de Y a fijar valores de las independientes. Hay entonces un objetivo fijado que es la determinación de una función de regresión. En ocasiones el interés del investigador es también analizar la variación de Y respecto a esta función lo que plantea un problema que es descrito por una distribución de probabilidad (Carlos N. Bouza, 2018).

En una regresión se busca una función $f(x)$ que pase por los puntos formados por la variable de salida y la variable de entrada. En la regresión, adicionalmente encontramos ruido, entendido como irregularidad indeseable en los datos, que se añade a los datos de salida de la función desconocida $y = f(x) + \xi$

X = variable controlada, predictor independiente, explicativa.

Y = variable respuesta, predicción, dependiente, regresando.

El ruido en la regresión se genera por las variables que tienen un efecto en los datos de salida, pero no se las conoce. El cálculo de la función se hace mediante de un método de estimación de parámetros poblacionales, como el de mínimos cuadrados ordinarios que busca la distancia más corta entre los valores reales y predichos (Alpaydin, 2014).

2.11 Librería SciPy

SciPy es una colección de algoritmos matemáticos y funciones de conveniencia construidas sobre la extensión Numpy de Python. Agrega un poder significativo a la sesión interactiva de Python al proporcionar al usuario comandos y clases de alto nivel para manipular y visualizar datos. Con SciPy, una sección interactiva de Python se

convierte en un entorno de procesamiento de datos y creación de prototipos de sistemas que rivaliza con sistemas como MATLAB, IDL, Octave, R-Lab y Scilab.

El beneficio adicional de usar SciPy en Python es que esto también hace que un poderoso lenguaje de programación esté disponible para usar en el desarrollo de programas sofisticados y aplicaciones especializadas. Las aplicaciones científicas que utilizan SciPy se benefician del desarrollo de módulos adicionales en numerosos nichos del panorama del software por parte de desarrolladores de todo el mundo. Todo, desde la programación paralela hasta las subrutinas y clases web y de bases de datos, se ha puesto a disposición del programador de Python. Todo este poder está disponible además de las bibliotecas matemáticas en SciPy (scipy.org, 2023).

2.12 Estadísticas (SciPy.stats)

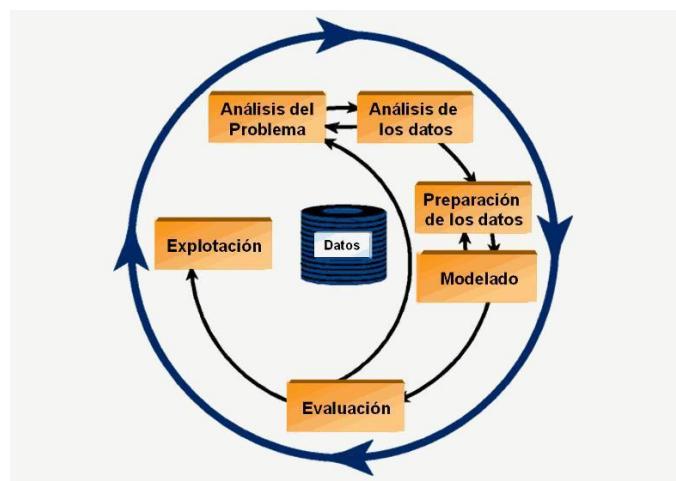
Todas las funciones de estadística se encuentran en el sub paquete SciPy.stats y se puede obtener una lista bastante completa de estas funciones. La lista de las variables aleatorias disponibles también se puede obtener de la cadena de documentación del sub paquete de estadísticas (scipy.org, 2023).

2.13 Metodología CRISP-DM

CRISP-DM, que significa Cross-Industry Standard Process for Data Mining, es una forma comprobada en la industria de guiar sus esfuerzos de minería de datos (CRISP-DM, 2021).

- Como **metodología**, incluye descripciones de las fases típicas de un proyecto, las tareas involucradas en cada fase y una explicación de las relaciones entre estas tareas (CRISP-DM, 2021).
- Como **modelo de proceso**, CRISP-DM proporciona una descripción general del ciclo de vida de la minería de datos (CRISP-DM, 2021).

Gráfico 16 El ciclo de vida de la minería de datos Autor: (CRISP-DM, 2021)



El modelo de ciclo de vida consta de seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre las fases según sea necesario (CRISP-DM, 2021).

El modelo CRISP-DM es flexible y se puede personalizar fácilmente. Por ejemplo, si su organización tiene como objetivo detectar el lavado de dinero, es probable que examine grandes cantidades de datos sin un objetivo de modelado específico. En lugar de modelar, su trabajo se centrará en la exploración y visualización de datos para descubrir patrones sospechosos en los datos financieros. CRISP-DM le permite crear un modelo de minería de datos que se ajuste a sus necesidades particulares (CRISP-DM, 2021).

En tal situación, las fases de modelado, evaluación e implementación pueden ser menos relevantes que las fases de comprensión y preparación de datos. Sin embargo, aún es importante considerar algunas de las preguntas planteadas durante estas últimas fases para la planificación a largo plazo y los objetivos futuros de extracción de datos (CRISP-DM, 2021).

CAPÍTULO III: METODOLOGÍA Y TÉCNICAS

3 Metodología y Técnicas

Utilizar la metodología CRISP DM para demostrar el uso de los algoritmos de Machine Learning seleccionados, conjuntamente con el dataset y su aplicación en la gestión de riesgos cuantitativos en proyectos predictivos y adaptativos.

En este capítulo se describen las acciones tomadas para la demostración de los algoritmos de machine learning seleccionados, conjuntamente el caso de uso de donde se origina el dataset y su aplicación en la gestión de riesgos cuantitativos en proyectos, mediante la aplicación del ciclo de vida que recomienda la metodología CRISP-DM.

3.1 Caso de uso

Para este trabajo de titulación usaremos el siguiente caso de uso:

Automatización del Nuevo Aeropuerto Internacional de Quito – Ecuador NQIA.

La automatización se da en la Construcción del Nuevo Aeropuerto Internacional de Quito.

Tuve el agrado de ser el Project Manager de la Automatización del nuevo Aeropuerto, por lo que, conozco de primera mano la problemática y como la gestión de riesgos cualitativos y cuantitativos nos ayudó para lograr cumplir con los niveles de calidad, presupuesto, y tiempo.

3.2 Comprensión del negocio

La fase inicial debe tomarse el tiempo para explorar lo que su organización espera obtener de la minería de datos, trate de involucrar a tantas personas clave como sea posible en estas discusiones y documente los resultados. El paso final de esta fase CRISP-DM analiza cómo producir un plan de proyecto utilizando la información recopilada en esta etapa (CRISP-DM, 2021).

3.2.1 Contrato de Concesión y Alianza Estratégica.

Corporación Quiport S.A, a través de la suscripción del Contrato de Concesión, es la concesionaria exclusiva del servicio aeroportuario de Quito, dentro de lo cual se

comprometió a operar y mantener el antiguo Aeropuerto de Quito ya desarrollar, construir, operar y mantener el Nuevo Aeropuerto Internacional Mariscal Sucre de Quito (Quiport, 2017).

El referido proceso de contratación se realizó bajo el auspicio del Gobierno de Canadá, representado por la Canadian Commercial Corporation (CCC) (Quiport, 2017).

Dentro de sus obligaciones contractuales, en los 35 años de concesión (desde 2006 hasta 2041), Corporación Quiport S.A. se compromete a realizar las inversiones necesarias para mantener el nivel de servicio contratado (Quiport, 2017).

Canadian Commercial Corporation contrata al Consorcio AECON - AG para la construcción del nuevo Aeropuerto Internacional de Quito NQIA, donde AECON de Canadá y AG Andrade Gutiérrez de Brasil (Quiport, 2017).

Consorcio AECON – AG contrata a la empresa Johnson Controls INC. para la automatización total del Nuevo Aeropuerto Internacional de Quito, con un presupuesto de 12 millones de dólares de Norte América.

Los sistemas a implementar son:

1. Sistema BAS Building Automation System
2. Sistema Contra Incendios - Fire Alarm
3. Sistema de Gestión Aeroportuaria
4. Sistema de control de accesos – seguridad electrónica
5. Sistema CCTV – seguridad electrónica
6. Sistema Plublic Adress
7. Sistema de iluminación de la pista de aterrizaje
8. Sistema HVAC (Climatización, Ventilación y Aire acondicionado)
9. Sistema de aproximación aérea
10. Sistema de aeronavegación
11. Sistema Lan
12. Sistema de VoIP
13. Sistema de Master Clock
14. Sistema de información AODB/FIDS/BIDS/GM/CUTE

Todos los sistemas incluyen diseño y equipo electrónico

3.2.2Determinar Objetivos del Negocio

3.2.2.1 Objetivo general

Determinar el análisis comparativo y conclusiones de la aplicación de algoritmos de machine Learning seleccionados en la aplicación de la gestión de riesgos cuantitativos en portafolio de proyectos involucrados en la automatización del nuevo aeropuerto internacional de Quito – NQIA.

3.2.2.2 Objetivos específicos

- Aplicar los siguientes algoritmos de machine learning:
 - Monte Carlo
 - Árboles de decisión
 - Regresión lineal
- Proyectos a recrear los algoritmos de machine learning seleccionados:
 - NQIA – ITT Systems
 - Sistema Lan
 - Sistema de VoIP
 - Sistema de Master Clock
 - Sistema de información AODB/FIDS/BIDS/GM/CUTE
 - Seguridad Electrónica
 - Sistema CCTV
 - Control de Accesos
- Realizar los siguientes análisis de riesgo cuantitativo:
 - Presupuesto - costo
 - Cronograma – tiempo
- Obtener un análisis comparativo de los resultados obtenidos al momento de aplicar los algoritmos de machine learning seleccionados.
- Determinar Conclusiones y Recomendaciones.

3.2.3Criterios de éxito del negocio

- Desarrollar el análisis comparativo de la aplicación de los algoritmos de machine learning seleccionados para poder determinar el nivel de efectividad de cada uno de ellos.
- Determinar el comportamiento de las estimaciones y predicciones tanto en presupuesto como en tiempo de los proyectos seleccionados
- Concluir que algoritmos aplican de mejor manera en el caso de uso seleccionado.

3.2.4Evaluar la situación

Para la recreación del caso de uso se usará un cronograma de actividades con recursos, tiempos, y fechas reales al momento de la construcción del Nuevo Aeropuerto Internacional de Quito.

El presupuesto a usar será modificado a propósito sus costos con el fin de garantizar la confidencialidad de la información económica original.

3.2.5Inventario de Recursos

- Recursos de hardware
 - 1 computador portátil
 - 1 impresora
 - 1 monitor a color de 20"
- Equipo de oficina
 - Acceso a Internet
 - Material de oficina
- Software
 - Ofimática
 - Microsoft Project Professional 2016
 - Herramienta colaborativa de Google
- Fuente de datos – datasets
 - Cronograma – planificación
 - Presupuestos

- Tablas de estimaciones tanto para cronograma, así como para el presupuesto.

3.2.6 Requisitos, supuestos y restricciones

- Requisitos
 - Contar con los datasets pertinentes para realizar las recreaciones planteadas
 - Contar con el computador apto para realizar las simulaciones requeridas
 - Exista secuencia lógica en los pasos a realizar
- Restricciones
 - Dataset de presupuesto debe ser alterado en sus costos originales con el afán de precautelar la información original.
 - Realizar cuadros de estimaciones como insumo para las simulaciones.
 - Realizar cuadros comparativos para comprobar los resultados obtenidos.

3.2.7 Riesgos y contingencias

Tabla 2 Riesgos y contingencias del proyecto. Autor: (Osejo Francisco, 2023)

RIESGOS	CONTINGENCIA
No entender los conceptos de las librerías de machine learning a usar.	Realizar la correspondiente investigación de los conceptos necesarios.
Cometer errores en la interpretación de los resultados obtenidos en los laboratorios.	Realizar la correspondiente investigación de los conceptos necesarios.
Errores en las estimaciones realizadas tanto en presupuesto como en tiempo del dataset a usar	Aplicar Juicio de expertos

3.2.8Costos

Tabla 3 Costos proyecto. Autor: (Osejo Francisco, 2023)

RUBRO	VALOR
Materiales	\$ 500
Subtotal Sin Iva:	\$ 500
Total general:	\$560

3.2.9Determinación de los objetivos de modelación

- Monte Carlo:
 - Realizar la simulación con el uso de librerías **SciPy.stats** y las requeridas.
 - Realizar iteraciones randomicas de 1.000, 5.000, 10.000 y 20.000
 - Realizar las gráficas necesarias y documentar los resultados comparativos.
- Regresión lineal simple y múltiple:
 - Realizar Cross Validation con el 80% de la data para entrenamiento y 20% de la data para prueba.
 - Seleccionar las variables independientes necesarias para pronosticar el costo – presupuesto.
 - Realizar el análisis de coeficientes
 - Realizar las gráficas necesarias y documentar los resultados comparativos
 - Conclusiones
- Árboles de decisión
 - Realizar Cross Validation con el 80% de la data para entrenamiento y 20% de la data para prueba.
 - Seleccionar las variables independientes necesarias para pronosticar el costo – presupuesto.
 - Realizar el análisis de coeficientes
 - Realizar las gráficas necesarias y documentar los resultados comparativos
 - Conclusiones

3.2.10 Plan de proyecto

Tabla 4 Plan del proyecto. Autor: (Osejo Francisco, 2023)

Tarea	Fecha de Inicio	Fecha de fin	Entregable
Comprendión del negocio	10 enero 2023	20 enero 2023	Autorización de datos
Compresión de los datos	21 enero 2023	15 febrero 2023	Entrega de datos
Preparación de los datos	16 febrero 2023	16 marzo 2023	Base de datos final
Modelado	17 marzo 2023	01 junio 2023	Modelo de datos
Implementación	02 junio 2023	31 julio 2023	Acta de entrega recepción

3.3 Comprensión de datos

La fase de comprensión de datos de CRISP-DM implica observar más de cerca los datos disponibles para la minería. Este paso es fundamental para evitar problemas inesperados durante la siguiente fase, la preparación de datos, que suele ser la parte más larga de un proyecto (CRISP-DM, 2021).

La compresión de los datos implica acceder a los datos y explorarlos mediante tablas y gráficos. Esto le permite determinar la calidad de los datos y describir los resultados de estos pasos en la documentación del proyecto (CRISP-DM, 2021).

3.3.1 Recopilación de datos iniciales

Los datos para utilizar en este proyecto provienen de dos fuentes:

- Cronograma – Tiempo. - Data correspondiente a la planificación de la implementación de la automatización de los siguientes contratos - sistemas:
 - Lan
 - VoIP
 - Master Clock

- AODB/FIDS/BIDS/GM/CUTE Designing
 - System Integration Implementation
 - Security
- Presupuesto – Costo. - Data correspondiente al presupuesto de la implementación de la automatización de los siguientes contratos – sistemas:
 - Lan
 - VoIP
 - Master Clock
 - AODB/FIDS/BIDS/GM/CUTE Designing
 - System Integration Implementation
 - Security

La información suministrada tanto en costo, así como en tiempo han sido modificados por aspectos de confidencialidad.

La información corresponde al período de 2011 a 2013 etapa en la cual se llevó a cabo la construcción del Nuevo Aeropuerto Internacional de Quito NQIA.

La información a usar en este trabajo de titulación corresponde a una parte de la automatización de nuevo aeropuerto, por lo que, no se considera información totalmente confiable.

3.3.2 Describiendo los datos

Tanto la data – información comprendida en el cronograma, así como del presupuesto se encuentra en sus formatos originales como son archivo en Microsoft Project para el caso del cronograma y para el presupuesto se encuentra en un archivo de Excel.

La información en cuestión debe pasar por un proceso de preparación de datos para que se convierta en datasets funcionales y óptimos.

3.3.2.1 Cronograma – tiempo. -

La información que contiene este archivo en Microsoft Project es la siguiente:

- 18 registros
- 7 columnas

- Código de WBS (Identificación de la actividad)
- Descripción de la actividad
- Fecha de inicio de la actividad
- Fecha de Fin de la actividad
- Estimación pesimista de la duración de la actividad en días
- Estimación más probable de la duración de la actividad en días
- Estimación optimista de la duración de la actividad en días

3.3.2.2 Presupuesto – costo. -

La información que contiene este archivo en Microsoft Excel es la siguiente:

- 21 registros
- 6 columnas
- Código de WBS (Identificación de la actividad)
- Descripción de la actividad
- Duración en días
- Estimación pesimista del costo de la actividad en dólares americanos
- Estimación más probable del costo de la actividad en dólares americanos
- Estimación optimista del costo de la actividad en dólares americanos

3.3.2.3 Estructura de los datos

3.3.2.3.1 Cronograma – tiempo

La información tiene la siguiente estructura de datos:

Tabla 5 Estructura de datos de dataset cronograma tiempo. Autor: (Osejo Francisco, 2023)

#	Campo	Descripción	Tipo de dato
1	WBS	Código de WBS (Identificación de la actividad)	Numérico
2	Descripción	Descripción de la actividad	String
3	Inicio	Fecha de inicio de la actividad	Fecha

4	Final	Fecha de Fin de la actividad	Fecha
5	Pesimista	Estimación pesimista de la duración de la actividad en días	Numérico
6	Mas_probable	Estimación más probable de la duración de la actividad en días	Numérico
7	Optimista	Estimación optimista de la duración de la actividad en días	Numérico

3.3.2.3.2 Presupuesto – costo

La información tiene la siguiente estructura de datos:

Tabla 6 Estructura de datos de dataset presupuesto costo. Autor: (Osejo Francisco, 2023)

#	Campo	Descripción	Tipo de dato
1	Descripción	Descripción de la actividad	String
2	Tiempo	Duración en días	Numérico
3	Recursos	Cantidad de recurso humano	Numérico
4	Pesimista	Estimación pesimista del costo de la actividad en dólares americanos	Numérico
5	Mas_probable	Estimación más probable del costo de la actividad en dólares americanos	Numérico
6	Optimista	Estimación optimista del costo de la actividad en dólares americanos	Numérico

3.3.3 Exploración de los datos

En el numeral anterior se representa una descripción de los datos, ahora se pasa a explorarlos mediante imágenes que permitan tener una idea inicial del comportamiento y contenido de los datos, de sus características, probablemente esto ayude a obtener segmentaciones de datos.

Gráfico 17 Cronograma – tiempo Autor: (Osejo Francisco, 2023)

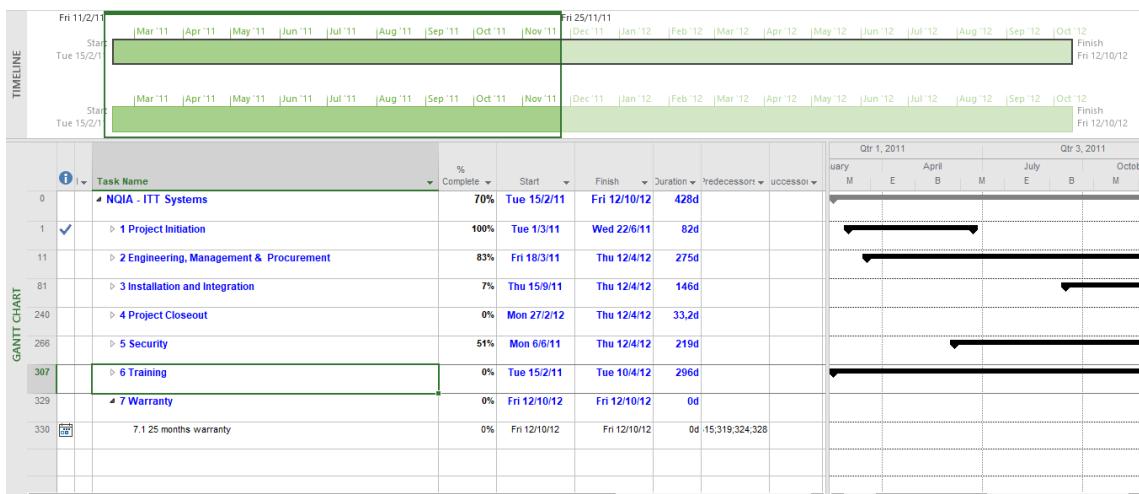


Gráfico 18 Presupuesto – costo Autor: (Osejo Francisco, 2023)

Customer	Contract	Contract Amount	Variation Order Claim	Others Variation Orders	Total	COST OBRA			% Ejecutado	Total Ejecutado	Planeado	Saldo	Total Real • Planeado
						Product	Service	Total					
Donoso BAS		183.608	0		183.608	-	106.504	106.504		-		36.806	69.698
						BAS JCI ->		106.504					
						Labor JCI ->		-					
						FMT JCI ->		-					
Aeon ITT	5.493.500	547.400	-190.693	5.850.207	5.850.207	3.429.930	1.283.050	4.712.980	90%	1.508.085	183.527	(7.441)	1.691.612
						FIDS/BIDS/RMIS/ADDB/CUTE - Lumpsum (Ultra) ->		1684.171					
						Servers y SD ->	32.411	32.411	100%	92.708	-	(297)	92.708
						Workstation (HP) ->	85.015	85.015	100%	91.683	-	(6.668)	91.583
						DDC (ACER) ->	78.000	78.000	80%	62.040	-	15.960	62.040
						Monitores 42" ->	103.500	103.500	102%	111.798	-	(2.298)	111.798
						Brackets for FIDS Monitors + Installation ->	50.000	50.000	90%	45.069	-	4.931	45.069
						LAN and VoIP - Materiales (Cisco / DOS) ->	1.370.713	1.370.713	100%	1.370.713	-	(0)	1.370.713
						LAN and VoIP - Servicios (Cisco / DOS) ->	208.950	208.950	100%	208.949	-	1	208.949
						Master Clock (Moba Time) ->	10.121	10.121	100%	18.182	-	(8.061)	18.182
						Ver Tab ITT COST DETAIL ->				200.748	-	#####	200.748
						Labor JCI ->	344.597	344.597	178%	614.466	50.959	#####	665.425
						FMT JCI ->	323.503	323.503	76%	249.418	-	80.085	249.418
						FMT (Travel / Hotel / Bonind etc.) ->							
						Warranty ->	50.000	50.000	0%	-	-	50.000	
						Risk ->	100.000	100.000	183%	189.310	-	(89.310)	189.310
						Contingency ->	200.000	200.000	0%	-	-	200.000	
Aeon Security	2.176.244	272.600			2.448.844	1.483.394	388.632	1.872.026	90%	1.335.143	148.349	(98)	1.483.492
						Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	1.483.394	1.483.394		-	-	75.000	
						Subcontract Installation ->	75.000	75.000					
						Ver Tab Security Cost Detail ->				337.963	-	#####	337.963
						Labor JCI ->	200.000	200.000	71%	141.261	-	58.739	141.261
						FMT JCI ->	115.632	115.632	0%	-	-	113.632	
Canala PA / FD / LC	1.150.000	0	-97.386	1.052.614	636.560	269.170	905.730			-	-	905.730	
						Public Address (Bosh) ->	380.000	380.000	95%	361.446	-	18.554	361.446
						Ingeniería PA (Bosh) ->	35.000	35.000	0%	-	-	35.000	
						Notifier ->	256.560	256.560	83%	219.924	-	44.636	219.924
						Watstopper ->	40.000	40.000	97%	38.752	-	1.248	38.752
						Labor JCI ->	114.578	114.578	45%	51.000	-	63.578	51.000
						FMT JCI ->	73.592	73.592	0%	-	-	79.592	
Total NQIA	9.003.352	820.000	-288.079	9.535.273	5.549.884	2.047.355	7.597.239	96%	7.310.355	382.835	809.779	7.693.189	

3.4 Preparación de datos

la preparación de datos es uno de los aspectos más importantes y, a menudo, más laboriosos de la minería de datos. De hecho, se estima que la preparación de datos suele ocupar entre el 50% y el 70% del tiempo y el esfuerzo de un proyecto. Dedicar la energía adecuada a las fases anteriores de comprensión del negocio y compresión de los datos puede minimizar esta sobrecarga, pero aún debe dedicar una buena cantidad de esfuerzo a preparar y empaquetar los datos para la minería (CRISP-DM, 2021).

Para el caso de este trabajo de titulación la información tanto del cronograma – tiempo y presupuesto – costo se encuentran aún en formatos estructurados tanto en Microsoft Project, así como en Microsoft Excel respectivamente.

Tal como están al momento no es posible trabajar con dicha información en los formatos en cuestión.

Por lo que, se necesita transformar la información de ambos al formato simple en Excel, inicialmente.

3.4.1 Cronograma – tiempo

En este caso se requiere trasladar la información del cronograma en formato Microsoft Project a formato en tipo Excel, manteniendo la misma estructura de datos descrita en la tabla 5 de este documento.

El dataset de cronograma – tiempo se muestra en el Gráfico 19.

Gráfico 19 Cronograma – tiempo en formato Excel Autor: (Osejo Francisco, 2023)

A	B	C	D	E	F	G
WBS	Descripción	Inicio	Final	Pesimista	Mas_probable	Optimista
0	NQIA - ITT Systems	15 February 2011 8:00	12 October 2012 8:00	435	428	420
1	Project Initiation	1 March 2011 8:00	22 June 2011 17:00	97	82	80
1.1	Mobilization	1 March 2011 8:00	22 June 2011 17:00	85	82	75
1.1.1	Notice to Proceed	1 March 2011 8:00	1 March 2011 8:00	3	0	0,57
1.1.2	Contract Signed	1 March 2011 8:00	11 April 2011 17:00	50	30	25
1.1.3	Performance Bond	12 April 2011 8:00	2 May 2011 17:00	30	15	10
1.1.4	Initial Office/Storage Trailer Setup	1 April 2011 8:00	22 June 2011 17:00	75	59	45
1.2	Project Planning	1 March 2011 8:00	18 May 2011 17:00	63	57	55
1.2.1	Project Planning & Scheduling Complete	1 March 2011 8:00	11 May 2011 17:00	75	52	45
1.2.2	Generate Project Management / Execution Plan Documents	1 March 2011 8:00	11 May 2011 17:00	65	52	50
1.2.3	Approval Planning Documents	12 May 2011 8:00	18 May 2011 17:00	7	5	4
2	Engineering, Management & Procurement	18 March 2011 8:00	12 April 2012 17:00	280	275	270
2.1	Monitoring and Control	18 March 2011 8:00	12 April 2012 17:00	281	275	271
2.1.1	Project Management Plan Updating	18 March 2011 8:00	12 April 2012 17:00	300	275	260
2.1.2	Reports	25 March 2011 8:00	12 April 2012 17:00	280	269	250
2.1.3	Control Changes Updating	25 March 2011 8:00	12 April 2012 17:00	300	270	250
2.1.4	Risk Matrix Updating	25 March 2011 8:00	12 April 2012 17:00	280	270	260
2.1.5	Schedule Report Updating	25 March 2011 8:00	12 April 2012 17:00	290	270	240
2.2	LAN Designing	19 May 2011 8:00	10 February 2012 15:00	190	186,75	180
2.2.1	Generate documentation for Preliminary Design Review (PDR) LAN - Technical Update	19 May 2011 8:00	16 July 2011 17:00	50	42	30
2.2.2	Customer PDR Comments LAN - Technical Update	2 June 2011 8:00	16 July 2011 17:00	40	32	28
2.2.3	PDR Approval	16 July 2011 8:00	16 July 2011 8:00	5	0	0,2
2.2.4	Generate documentation for Critical Design Review (CDR) LAN and I&N	11 July 2011 8:00	11 January 2012 16:36	65	59	45

3.4.2Presupuesto – costo

En este caso el formato del presupuesto ya se encuentra en formato Excel, sin embargo, existe la necesidad de modificar su información aún más para que sea tratada en las siguientes etapas.

Gráfico 20 Presupuesto – costo en formato Excel Autor: (Osejo Francisco, 2023)

A	B	C	D	E	F
Descripcion	Tiempo	Recursos	Optimista	Mas_probable	Pesimista
FIDS/BIDS/RMS/AODB/CUTE - Lumpsum (Ultra) ->	292,88	20	982070,48	984171	986271,52
Servers y SO ->	27	5	78549,02	92410,61	104423,99
Workstation (HP) ->	45	10	78213,80	85015	96917,10
DDC (ACER) ->	72	10	73320	78000	92040
Monitores 42->	71	10	96360	109500	122640
Brackets for FIDS Monitors + Imstallation ->	65	10	44500	50000	58000
LAN and VoIP - Materiales (Cisco / DOS) ->	194	5	974763,16	975713,08	983734,30
LAN and VoIP - Servicios (Cisco / DOS) ->	444,63	15	190144,05	208949,50	231933,95
Master Clock (Moba Time) ->	230	2	9310,95	10120,60	11638,69
FIDS Labor JCI ->	188,88	10	323921,36	344597,19	406624,68
FIDS FMT JCI ->	64	5	286667,62	329503,01	382223,49
Security Systems: Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	287	20	975718,44	983393,53	995238,30
Subcontract Installation ->	89	15	66750	75000	85500
Security Labor JCI ->	104	20	190000	200000	226000
Security FMT JCI ->	50	5	106814,19	113632,12	123859,01
Public Address (Bosh) ->	90	10	353400	380000	399000
Ingenieria PA (Bosh) ->	30	10	31850	35000	39200
Notifier ->	120	5	246297,60	256560	289912,80
Wattstopper ->	180	5	37600	40000	45600
PA Labor JCI ->	110	10	105411,39	114577,60	127181,14
PA FMT JCI ->	90	10	72428,72	79592	94714,48

3.4.3Selección de datos

En este punto, se está listo para comenzar a seleccionar los datos relevantes para sus objetivos de minería de datos. En general, hay dos formas de seleccionar datos: (CRISP-DM, 2021)

- Seleccionar elementos (filas) implica tomar decisiones como qué cuentas, productos o clientes incluir (CRISP-DM, 2021).
- La selección de atributos o características (columnas) implica tomar decisiones sobre el uso de características como el monto de la transacción o los ingresos del hogar (CRISP-DM, 2021).

Para el caso del presente trabajo de titulación contamos con dos datasets estructurados y bien definidos.

3.4.4 Cronograma – tiempo

La gestión de riesgos cuantitativos lo podemos visualizar en el Gráfico 6, es importante recalcar que el análisis cuantitativo – numérico se lo realiza en aquellos riesgos identificados y que ameriten de este análisis. Es decir, no es mandatorio aplicar a todo el set de datos, en este caso a todo el cronograma del proyecto, ya que, las actividades o grupos de actividades y sus duraciones – tiempo se las ejecuta en determinado momento y/o en paralelo, así como la presencia de dependencias entre actividades. Por lo tanto, no puede ser lineal.

En conclusión, existe la necesidad de seleccionar un grupo de actividades que amerite realizar el análisis cuantitativo – numérico.

Para el caso de este trabajo de titulación, nos enfocaremos en el siguiente grupo de actividades:

Instalación del subsistema AODB / FIDS / BIDS / GM / CUTE. - Este corresponde al subsistema de información aeroportuaria, es decir, a las pantallas – monitores informativos de arribos, despegue, retrasos. Este subsistema es muy importante y crucial para obtener las certificaciones de operación aeroportuaria, sin este subsistema el aeropuerto no puede iniciar operaciones.

Gráfico 21 FID – Monitor informativo arribos y salidas Autor: (Osejo Francisco, 2023)



Gráfico 22 FID – monitor curvo informativo arribos y salidas Autor: (Osejo Francisco, 2023)



Los datos seleccionados para el dataset Cronograma – tiempo son:

Tabla 7 Datos seleccionados dataset cronograma - tiempo. Autor: (Osejo Francisco, 2023)

Variable	Identificación	Tipo de variable
Descripción	Descripción de la actividad	Object
Pesimista	Estimación pesimista de la duración de la actividad en días	float
Mas_probable	Estimación más probable de la duración de la actividad en días	float
Optimista	Estimación optimista de la duración de la actividad en días	float

3.4.5 Presupuesto – costo

Los datos seleccionados para el dataset Presupuesto – costo son:

Tabla 8 Datos seleccionados dataset presupuesto - costo. Autor: (Osejo Francisco, 2023)

Variable	Identificación	Tipo de variable
Descripcion	Descripción de la actividad	Object
Tiempo	Duración en días	float
Recursos	Cantidad de recurso humano	float
Pesimista	Estimación pesimista del costo de la actividad en días	float
Mas_probable	Estimación más probable del costo de la actividad en días	float
Optimista	Estimación optimista del costo de la actividad en días	float

3.4.6 Limpieza de datos

Limpiar sus datos implica observar más de cerca los problemas en los datos que ha elegido incluir para el análisis (CRISP-DM, 2021).

Para el caso de este trabajo de titulación los datasets:

- Cronograma – tiempo
- Presupuesto – costo

Ambos tienen origen de datos estructurados y acondicionados para un fin específico, por lo que, la limpieza de datos se limita las siguientes actividades:

- Al momento de convertir los archivos de formato Excel a formato .CSV
 - CSV UTF 8 Delimitado por comas.
 - Reemplazar las comas (separador de decimales) por puntos
 - Reemplazar las punto y coma – delimitaciones de campos (;) por comas (,)

Ambos datasets no tienen muchos registros, por lo tanto, se puede realizar la limpieza de datos de forma manual.

Adicionalmente no existen campos con datos nulos y/o ambiguos

3.4.7 Construcción de nuevos datos

Es frecuente que necesite construir nuevos datos. Por ejemplo, puede ser útil crear una nueva columna que marque la compra de una garantía extendida para cada transacción (CRISP-DM, 2021).

Hay dos formas de construir nuevos datos:

- Derivación de atributos (columnas o características) (CRISP-DM, 2021).
- Generación de registros (filas) (CRISP-DM, 2021).

Para este trabajo de titulación existe la necesidad aplicar distribuciones de probabilidad, especialmente al momento de modelar la Simulación de Monte Carlos.

Las distribuciones de probabilidad a aplicar son:

- Distribución triangular
 - $tE = (tO + tM + tP) / 3$
 - $cE = (cO + cM + cP) / 3$
- Distribución Beta Pert.

- $tE = (tO + 4tM + tP)/6$
- $cE = (cO + 4cM + cP)/6$

Ambas distribuciones de probabilidad hay que aplicar en ambos datasets, el cálculo se las realizará en la hoja de cálculo en Excel, por facilidad y comprobación.

Gráfico 23 Cronograma – tiempo con nuevas variables Autor: (Osejo Francisco, 2023)

A	B	C	D	E	F	G	H	I	
1	WBS	Descripción	Inicio	Final	Pesimista	Mas_probable	Optimista	Distribucion_triangular	Distribucion_beta
2	3.6.4	AODB / FIDS / BIDS / GM / CUTE Installation	16 January 2012 8:00	12 April 2012 17:00	228,25	204,00	181,62	204,62	204,31
3	3.6.4.1	Installation and Configuration	16 January 2012 8:00	12 April 2012 17:00	67,00	64,00	61,00	64,00	64,00
4	3.6.4.2	Install and configure servers MCR1	29 February 2012 8:00	5 March 2012 17:00	6,00	4,00	3,00	4,33	4,17
5	3.6.4.3	Install and configure servers MCR2	1 March 2012 8:00	6 March 2012 17:00	6,00	4,00	2,00	4,00	4,00
6	3.6.4.4	Update system configurations	7 March 2012 8:00	20 March 2012 17:00	5,00	10,00	8,00	7,67	8,83
7	3.6.4.5	Airlines systems/info Ready	15 February 2012 8:00	15 February 2012 8:00	1,00	0,00	0,25	0,42	0,21
8	3.6.4.6	Configure airline interfaces	21 March 2012 8:00	12 April 2012 17:00	18,00	17,00	15,00	16,67	16,83
9	3.6.4.7	Install Displays Mountings/Brackets	16 January 2012 8:00	10 February 2012 17:00	22,00	20,00	18,00	20,00	20,00
10	3.6.4.8	Install LCDs & confirm operational	8 March 2012 9:36	29 March 2012 9:36	17,00	15,00	14,00	15,33	15,17
11	3.6.4.9	Configure DDC	24 February 2012 9:36	23 March 2012 9:36	22,00	20,00	18,00	20,00	20,00
12	3.6.4.10	Install DDC	9 March 2012 9:36	23 March 2012 9:36	12,00	10,00	9,00	10,33	10,17
13	3.6.4.11	Integration of DDC with Monitor and servers	12 March 2012 9:36	26 March 2012 9:36	13,00	10,00	8,00	10,33	10,17
14	3.6.4.12	Monitors and DDC are done	26 March 2012 9:36	26 March 2012 9:36	1,00	0,00	0,02	0,34	0,17
15	3.6.4.13	SiAT for FIDS	2 April 2012 9:36	12 April 2012 9:36	9,00	8,00	7,00	8,00	8,00
16	3.6.4.14	set up of CUSE	1 March 2012 8:00	14 March 2012 17:00	14,00	10,00	9,00	11,00	10,50
17	3.6.4.15	Install CUSE peripherals	6 March 2012 8:00	19 March 2012 17:00	12,00	10,00	8,00	10,00	10,00
18	3.6.4.16	Integration CUSE at locations	7 March 2012 8:00	7 March 2012 17:00	1,25	1,00	0,57	0,94	0,97
19	3.6.4.17	Site Acceptance Test (SiAT) for CUSE	12 March 2012 8:00	12 March 2012 17:00	2,00	1,00	0,78	1,26	1,13
20									

Tabla 9 Nuevas variables dataset cronograma - tiempo. Autor: (Osejo Francisco, 2023)

Variable	Identificación	Tipo de variable
Descripción	Descripción de la actividad	Object
Pesimista	Estimación pesimista de la duración de la actividad en días	float
Mas_probable	Estimación más probable de la duración de la actividad en días	float
Optimista	Estimación optimista de la duración de la actividad en días	float
Distribucion_triangular	Cálculo de distribución triangular	float
Distribucion_beta	Cálculo de distribución beta Pert	float

Gráfico 24 Presupuesto – costo con nuevas variables Autor: (Osejo Francisco, 2023)

Descripcion	Tiempo	Recursos	Optimista	Mas_probable	Pesimista	Distribucion_Triangular	Distribucion_beta
FIDS/BIDS/RMS/AODB/CUTE - Lumpsum (Ultra) ->	292,88	20	982070,48	984171	986271,52	984171	984171
Servers y SO->	27	5	78549,02	92410,61	104423,99	91794,54	92102,57
Workstation (HP) ->	45	10	78213,80	85015	96917,10	86715,30	85865,15
DDC (ACER) ->	72	10	73320	78000	92040	81120	79560
Monitores 42->	71	10	96360	109500	122640	109500	109500
Brackets for FIDS Monitors + Installation ->	65	10	44500	50000	58000	50833,33	50416,67
LAN and VoIP - Materiales (Cisco / DOS) ->	194	5	974763,16	975713,08	983734,30	978070,18	976891,63
LAN and VoIP - Servicios (Cisco / DOS) ->	444,63	15	190144,05	208949,50	231933,95	210342,50	209646,00
Master Clock (Moba Time) ->	230	2	9310,95	10120,60	11638,69	10356,75	10238,67
FIDS Labor JCI ->	188,88	10	323921,36	344597,19	406624,68	358381,08	351489,13
FIDS FMT JCI ->	64	5	286667,62	329503,01	382223,49	332798,04	331150,53
Security Systems: Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	287	20	975718,44	983393,53	995238,30	984783,42	984088,48
Subcontractor Installation ->	89	15	66750	75000	85500	75750	75375
Security Labor JCI ->	104	20	190000	200000	226000	205333,33	202666,67
Security FMT JCI ->	50	5	106814,19	113632,12	123859,01	114768,44	114200,28
Public Address (Bosh) ->	90	10	353400	380000	399000	377466,67	378733,33
Ingenieria PA (Bosh) ->	30	10	31850	35000	39200	35350	35175
Notifier ->	120	5	246297,60	256560	289912,80	264256,80	260408,40
Wattstopper ->	180	5	37600	40000	45600	41066,67	40533,33
PA Labor JCI ->	110	10	105411,39	114577,60	127181,14	115723,38	115150,49
PA FMT JCI ->	90	10	72428,72	79592	94714,48	82245,07	80918,53

Tabla 10 Nuevas variables dataset presupuesto - costo. Autor: (Osejo Francisco, 2023)

Variable	Identificación	Tipo de variable
Descripcion	Descripción de la actividad	Object
Tiempo	Duración en días	float
Recursos	Cantidad de recurso humano	float
Pesimista	Estimación pesimista del costo de la actividad en días	float
Mas_probable	Estimación más probable del costo de la actividad en días	float
Optimista	Estimación optimista del costo de la actividad en días	float
Distribucion_triangular	Cálculo de distribución triangular	float
Distribucion_beta	Cálculo de distribución beta Pert	float

3.5 Modelado

El modelado generalmente se lleva a cabo en múltiples iteraciones. Por lo general, los mineros de datos ejecutan varios modelos utilizando los parámetros predeterminados y luego ajustan los parámetros o vuelven a la fase de preparación de datos para las manipulaciones requeridas por su modelo de elección. Es raro que la pregunta de

minería de datos de una organización se responda satisfactoriamente con un solo modelo y una sola ejecución. Esto es lo que hace que la minería de datos sea tan interesante (CRISP-DM, 2021).

3.5.1 Selección de técnicas de modelado

De acuerdo al literal 2.7.3 de este documento, los algoritmos seleccionados para modelar son los siguientes:

- d) Monte Carlo
- e) Análisis de regresión
- f) Árboles de decisión

3.5.2 Construyendo los modelos

3.5.2.1 Simulación Monte Carlo

La simulación de Monte Carlo se explica con mayor detalle en la sección 2.8 de este trabajo de titulación.

Este modelo tiene por objetivo realizar lo siguiente:

- a) Aplicar Distribución Probabilística Triangular, ver literal 2.4.3 de este documento.
- b) Aplicar Distribución Beta o Pert, ver literal 2.4.5 de este documento.
- c) Usar los datasets de cronograma – tiempo y presupuesto – costo de forma separada para aplicar tanto la distribución triangular, así como la distribución Beta o Pert respectivamente.
- d) Aplicar la simulación de Monte Carlo haciendo uso de la librería SciPy.stats ver literal 2.11 de este documento para mayor explicación.
- e) Realizar la simulación de Monte Carlo usando los dataset de cronograma – tiempo y presupuesto – costo por separado aplicando de forma individual las distribuciones triangular y Beta – Pert.
- f) Aplicar la simulación de Monte Carlo para 1.000, 5.000, 10.000 y 20.000 iteraciones de manera independiente con cada dataset y con cada distribución probabilística seleccionada.
- g) Documentar el código en Python
- h) Graficar

- i) Registrar los resultados.
- j) Conclusiones

3.5.2.1.1 Aplicar Monte Carlo a Dataset de Cronograma – Tiempo

Gráfico 25 Simulación monte carlo – contenido dataset cronograma – tiempo Autor: (Osejo Francisco, 2023)

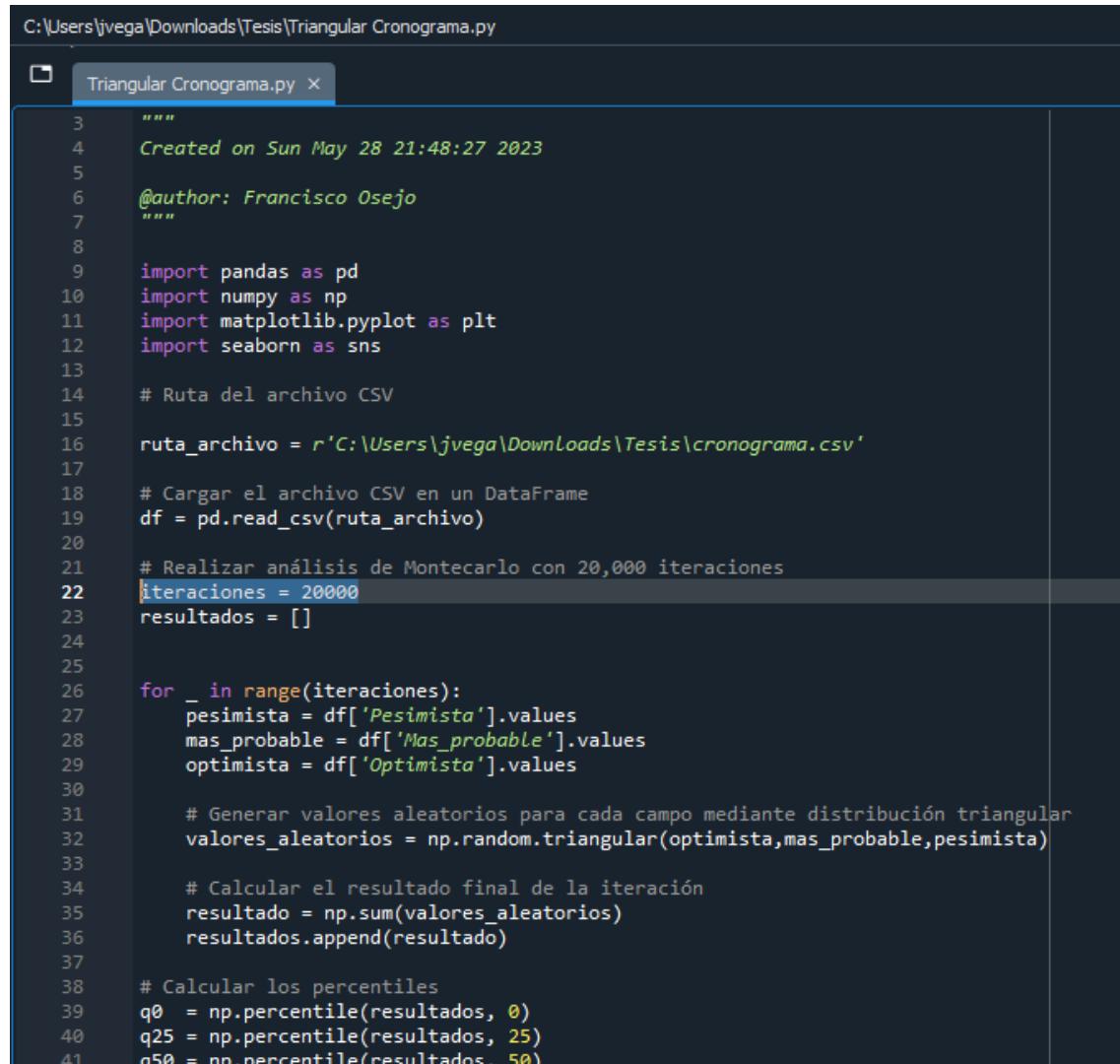
df - DataFrame

Index	WBS	Descripción	Inicio	Final	Optimista	Is probable	Pesimista	Ución trian	Probabilidad
0	3.6.4.1	Installation and Configuration	16 January 2012 8:00	12 April 2012 17:00	61	64	67	64	63.5
1	3.6.4.2	Install and configurate servers MCR1	29 February 2012 8:00	5 March 2012 17:00	3	4	6	4.33	3.89
2	3.6.4.3	Install and configurate servers MCR2	1 March 2012 8:00	6 March 2012 17:00	2	4	6	4	3.67
3	3.6.4.4	Update system configurations	7 March 2012 8:00	20 March 2012 17:00	8	10	12	10	9.67
4	3.6.4.5	Airlines systems/infos. Ready	15 February 2012 8:00	15 February 2012 8:00	1.5	2	3.5	2.33	1.97
5	3.6.4.6	Configure airline interfaces	21 March 2012 8:00	12 April 2012 17:00	15	17	18	16.67	16.61
6	3.6.4.7	Install Display's Mountings/Brackets	16 January 2012 8:00	10 February 2012 17:00	18	20	22	20	19.67
7	3.6.4.8	Install LCDs & confirm operational	8 March 2012 9:36	29 March 2012 9:36	14	15	17	15.33	14.89
8	3.6.4.9	Configure DDC	24 February 2012 9:36	23 March 2012 9:36	18	20	22	20	19.67
9	3.6.4.10	Install DDC	9 March 2012 9:36	23 March 2012 9:36	9	10	12	10.33	9.89
10	3.6.4.11	Integration of DDC with Monitor and servers	12 March 2012 9:36	26 March 2012 9:36	8	10	13	10.33	9.72
11	3.6.4.12	Monitors and DDC are done	26 March 2012 9:36	26 March 2012 9:36	12	15	17	14.67	14.44
12	3.6.4.13	SiAT for FIDS	2 April 2012 9:36	12 April 2012 9:36	7	8	9	8	7.83
13	3.6.4.14	set up of CUSE	1 March 2012 8:00	14 March 2012 17:00	9	10	14	11	10
14	3.6.4.15	Install CUSE peripherals	6 March 2012 8:00	19 March 2012 17:00	8	10	12	10	9.67
15	3.6.4.16	Integration CUSE at locations	7 March 2012 8:00	7 March 2012 17:00	13	15	16	14.67	14.61
16	3.6.4.17	Site Acceptance Test (SiAT) for CUSE	12 March 2012 8:00	12 March 2012 17:00	0.78	1	2	1.26	1.01

3.5.2.1.1.1 Distribución Probabilística Triangular

3.5.2.1.1.1.1 Código en Python Distribución Probabilística Triangular

Gráfico 26 Simulación monte carlo - código en Python distribución probabilística Triangular Cronograma – tiempo Autor: (Osejo Francisco, 2023)



The screenshot shows a code editor window with the file 'Triangular Cronograma.py' open. The code is written in Python and performs the following steps:

- Imports pandas, numpy, matplotlib.pyplot, and seaborn.
- Specifies the CSV file path: 'C:\Users\jvega\Downloads\Tesis\cronograma.csv'.
- Creates a DataFrame 'df' from the CSV file.
- Performs Monte Carlo analysis with 20,000 iterations.
- For each iteration:
 - Extracts values for 'Pesimista', 'Mas_probable', and 'Optimista' from the DataFrame.
 - Generates random numbers for each field using np.random.triangular.
 - Calculates the sum of these values to get the final result for the iteration.
- Calculates percentiles for the results: q0, q25, and q50.

```
C:\Users\jvega\Downloads\Tesis\Triangular Cronograma.py
Triangular Cronograma.py ×

3     """
4     Created on Sun May 28 21:48:27 2023
5
6     @author: Francisco Osejo
7     """
8
9     import pandas as pd
10    import numpy as np
11    import matplotlib.pyplot as plt
12    import seaborn as sns
13
14    # Ruta del archivo CSV
15
16    ruta_archivo = r'C:\Users\jvega\Downloads\Tesis\cronograma.csv'
17
18    # Cargar el archivo CSV en un DataFrame
19    df = pd.read_csv(ruta_archivo)
20
21    # Realizar análisis de Montecarlo con 20,000 iteraciones
22    iteraciones = 20000
23    resultados = []
24
25
26    for _ in range(iteraciones):
27        pesimista = df['Pesimista'].values
28        mas_probable = df['Mas_probable'].values
29        optimista = df['Optimista'].values
30
31        # Generar valores aleatorios para cada campo mediante distribución triangular
32        valores_aleatorios = np.random.triangular(optimista,mas_probable,pesimista)
33
34        # Calcular el resultado final de la iteración
35        resultado = np.sum(valores_aleatorios)
36        resultados.append(resultado)
37
38    # Calcular los percentiles
39    q0 = np.percentile(resultados, 0)
40    q25 = np.percentile(resultados, 25)
41    q50 = np.percentile(resultados, 50)
```

Gráfico 27 Simulación monte carlo - código en Python distribución probabilística triangular cronograma – tiempo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\Triangular Cronograma.py
  □ Triangular Cronograma.py ×
41 q50 = np.percentile(resultados, 50)
42 q75 = np.percentile(resultados, 75)
43 q95 = np.percentile(resultados, 95)
44 q100 = np.percentile(resultados, 100)
45
46 # Imprimir los resultados
47 print("Q25:", q25)
48 print("Q50 (Mediana):", q50)
49 print("Q75:", q75)
50 print("Q95:", q95)
51 print("Q100:", q100)
52
53 # Crear histograma
54 plt.figure(figsize=(10, 6))
55 sns.histplot(resultados, kde=True)
56 # Agregar líneas verticales en los percentiles
57 plt.axvline(q0, color='g', linestyle='--', label='Q0')
58 plt.axvline(q25, color='r', linestyle='--', label='Q25')
59 plt.axvline(q50, color='g', linestyle='--', label='Q50')
60 plt.axvline(q75, color='b', linestyle='--', label='Q75')
61 plt.axvline(q95, color='m', linestyle='--', label='Q95')
62 #plt.axvline(q100, color='g', linestyle='--', label='Q100')
63
64 # Agregar anotaciones con los valores de los percentiles en negrita y color negro
65 plt.text(q0, 100, f'Q0: {q0:.2f}', rotation=0, va='bottom', color='black', weight='bold')
66 plt.text(q25, 200, f'Q25: {q25:.2f}', rotation=0, va='bottom', color='black', weight='bold')
67 plt.text(q50, 600, f'Q50: {q50:.2f}', rotation=0, va='bottom', color='black', weight='bold')
68 plt.text(q75, 300, f'Q75: {q75:.2f}', rotation=0, va='bottom', color='black', weight='bold')
69 plt.text(q95, 50, f'Q95: {q95:.2f}', rotation=0, va='bottom', color='black', weight='bold')
70 #plt.text(q100, 0, f'Q100: {q100:.2f}', rotation=0, va='bottom', color='black', weight='bold')
71
72 plt.xlabel('Tiempo en número de días')
73 plt.ylabel('Frecuencia')
74 plt.title('Histograma Distribución Triangular - Aplicando Monte Carlo (Veinte Mil Iteraciones) - Cronograma - Tiempo')
75 plt.legend()
76 plt.show()
77
```

Gráfico 28 Simulación monte carlo - código en Python distribución probabilística triangular cronograma – tiempo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\Triangular Cronograma.py
```

```
Triangular Cronograma.py* x
```

```
78 # Graficar el histograma
79 plt.figure(figsize=(10, 6))
80 sns.histplot(resultados, kde=True)
81
82 # Rellenar áreas con diferentes colores
83 plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q0, q25, alpha=0.3, color='red', label='Q25')
84 plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q25, q75, alpha=0.3, color='green', label='Q50')
85 plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q75, q95, alpha=0.3, color='blue', label='Q75')
86 plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q95, q100, alpha=0.3, color='magenta', label='Q95')
87
88 # Agregar líneas verticales en los percentiles
89 plt.axvline(q25, color='r', linestyle='--')
90 plt.axvline(q50, color='g', linestyle='--')
91 plt.axvline(q75, color='b', linestyle='--')
92 plt.axvline(q95, color='m', linestyle='--')
93
94 # Agregar anotaciones con los valores de los percentiles en negrita y color negro
95 plt.text(q25, 200, f'Q25: {q25:.2f}', rotation=0, va='bottom', color='black', weight='bold')
96 plt.text(q50, 600, f'Q50: {q50:.2f}', rotation=0, va='bottom', color='black', weight='bold')
97 plt.text(q75, 300, f'Q75: {q75:.2f}', rotation=0, va='bottom', color='black', weight='bold')
98 plt.text(q95, 0, f'Q95: {q95:.2f}', rotation=0, va='bottom', color='black', weight='bold')
99
100 plt.xlabel('Tiempo en número de días')
101 plt.ylabel('Frecuencia')
102 plt.title('Histograma Distribución Triangular - Aplicando Monte Carlo (Veinte Mil Iteraciones) - Cronograma - Tiempo')
103 plt.legend()
104 plt.show()
105
106 # Calcular la media y la desviación estándar de los resultados
107 media = np.mean(resultados)
108 desviacion_estandar = np.std(resultados)
109
110 # Calcular la probabilidad de cumplir el presupuesto objetivo
111 # Presupuesto objetivo
112 presupuesto_objetivo = 250.027
113 probabilidad_cumplir = sum(resultado <= presupuesto_objetivo for resultado in resultados) / iteraciones
114
115 print("Probabilidad de cumplir el presupuesto objetivo:", probabilidad_cumplir)
116
```

Se realizó la simulación de Monte Carlo para 1.000, 5.000, 10.000 y 20.000 iteraciones aplicando **np.random.triangular ()** de **numpy** de manera independiente.

Generar valores aleatorios para cada campo mediante distribución triangular

valores_aleatorios = np.random.triangular(optimista,mas_probable,pesimista)
(Osejo Francisco, 2023)

Para efectos de este trabajo de titulación sólo mostraremos el código aplicado a las 20,000 iteraciones.

Para las 1.000, 5.000, 10.000 iteraciones hay que cambiar en el código el valor de la variable “**iteraciones =**”, **adicionalmente cambiar posiciones y textos en el código correspondiente a la generación de los histogramas respectivos en cada ejecución del programa.** (Osejo Francisco, 2023)

3.5.2.1.1.1.2 Histogramas obtenidos en Python Distribución Probabilística Triangular

Para efectos de este trabajo de titulación solo mostraremos Histogramas obtenidos en las simulaciones de 10,000 y 20,000 iteraciones

Gráfico 29 Histograma distribución triangular – aplicando monte carlo (diez mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)

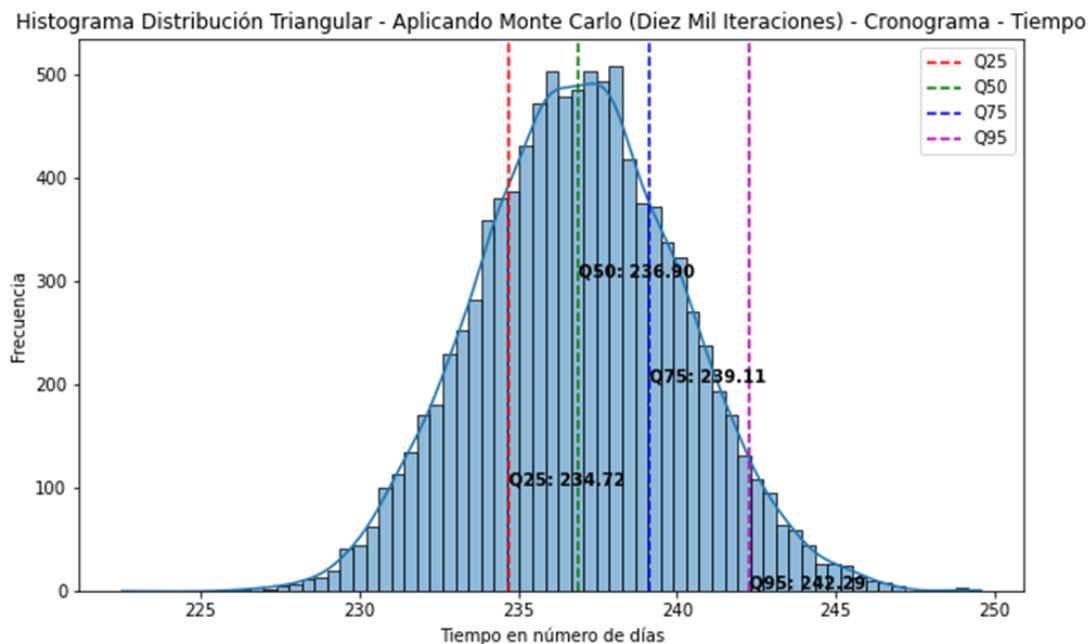
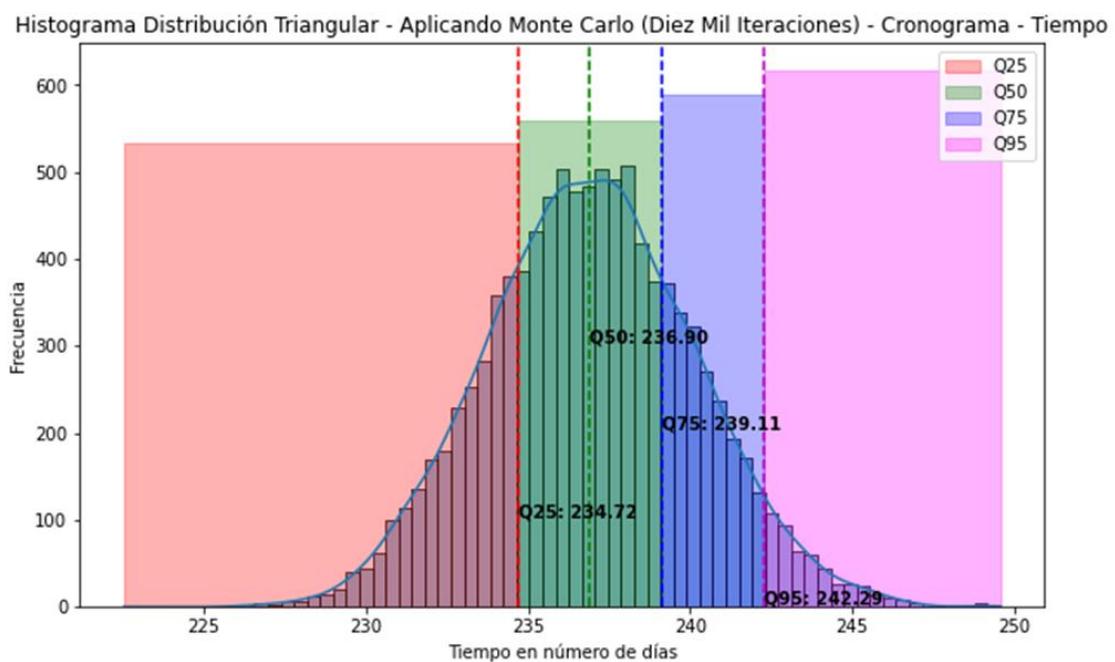


Gráfico 30 Histograma distribución triangular – aplicando monte carlo (diez mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)



Como podemos observar en los Gráficos 29 y 30 de este documento, el código del algoritmo realiza 10,000 diez mil iteraciones en donde aplicamos la fórmula de la Distribución triangular ($tE = (tO + tM + tP) / 3$), obteniendo la media de la sumatoria de los valores (optimista + más probable + pesimista) / 3, la librería de numpy `np.random.triangular(optimista,mas_probable,pesimista)` realiza los cálculos en base los valores generados randomicamente posibles dentro de los rangos de los valores de las variables optimista y pesimista (Osejo Francisco, 2023).

Obtenemos los percentiles Q0, Q25, Q50, Q75, Q95, Q100:

Tabla 11 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000 y 10,000 Autor: (Osejo Francisco, 2023)

Distribución Triangular	Simulación Monte Carlo - Iteraciones		
	1.000	5.000	10.000
Q0	226,884	225,758	222,516
Q25	234,720	234,790	234,712
Q50	236,840	237,010	236,866
Q75	239,050	239,200	239,084
Q95	241,960	242,330	242,273

Q100	248,371	249,386	249,582
------	---------	---------	---------

Q0: Muestra el valor mínimo en días que pueden iniciar las actividades

Q50: Muestra el valor medio en días en los cuales las actividades pueden ejecutarse

Q100: Muestra el valor máximo en días que pueden terminar las actividades

La interpretación de los Gráficos 29 y 30 es la siguiente:

- Permite identificar visualmente los límites de los rangos de tiempos en días
- Facilita el análisis de estimaciones de tiempo en días
- Muestra con mayor precisión las probabilidades de lograr el objetivo
 - Q0: Cualquier duración menor o igual al Q0 tendrá 0% de probabilidades de lograr el objetivo.
 - Q25: Cualquier duración menor o igual al Q25 tendrá de 0% a 25% de probabilidades de lograr el objetivo.
 - Q50: Cualquier duración menor o igual al Q50 tendrá de 26% a 50% de probabilidades de lograr el objetivo.
 - Q75: Cualquier duración menor o igual al Q75 tendrá de 51% a 75% de probabilidades de lograr el objetivo.
 - Q100: Cualquier duración menor o igual al Q100 tendrá de 76% a 100% de probabilidades de lograr el objetivo.
- Obtener los percentiles nos permite determinar el nivel de riesgo existente en el cumplimiento de los objetivos.
 - Q25: Indica que existe un 25% de probabilidad de éxito, lo que implica un alto riesgo de lograr el objetivo.
 - Q50: Indica que existe un 50% de probabilidad de éxito, lo que implica un riesgo medio de lograr el objetivo.
 - Q75: Indica que existe un 75% de probabilidad de éxito, lo que implica un riesgo bajo de lograr el objetivo.
 - Q100: Indica que existe un 100% de probabilidad de éxito, lo que implica que no existe riesgo de lograr el objetivo.

- Conseguir las ponderaciones de riesgo (Alto, medio, bajo) conjuntamente con las probabilidades de éxito permite determinar la matriz de probabilidad e impacto con esquema de puntuación que se muestra en el Gráfico 5
- La Frecuencia muestra claramente el nivel de iteraciones que coincidieron en ese rango de análisis. Por lo que, es muy importante revisar los resultados obtenidos en cada conjunto de iteraciones.
- La información descrita en la Tabla 11 muestra que a mayor número de iteraciones (1000, 5000, 10000) se logra afinar la precisión en los resultados obtenidos.

Gráfico 31 Histograma distribución triangular – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)

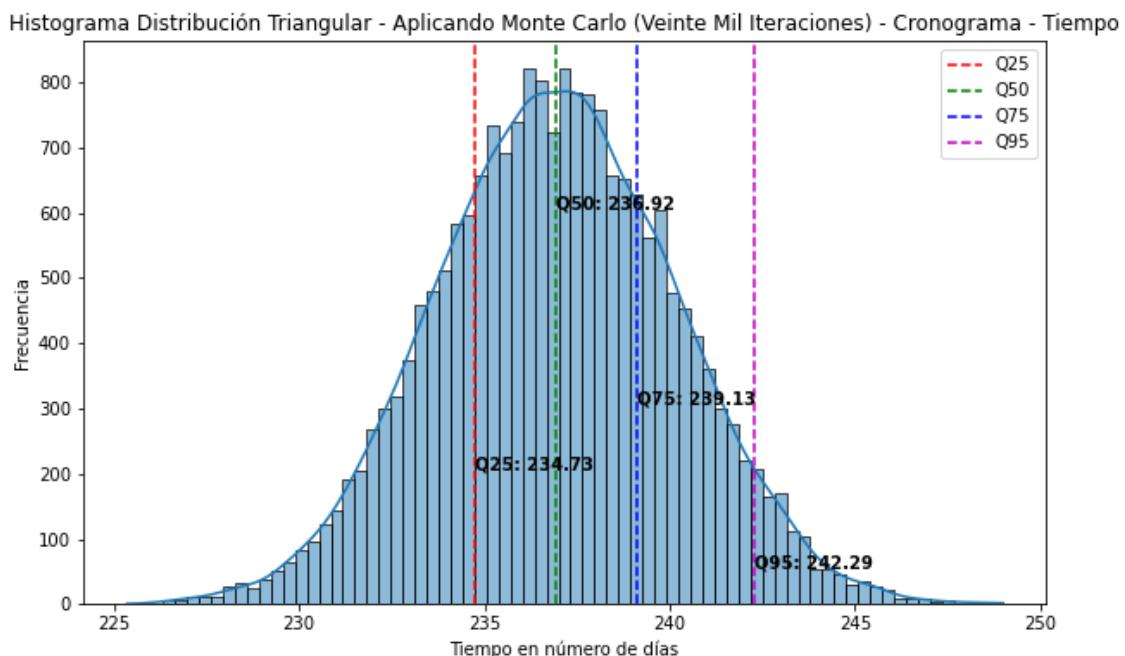


Gráfico 32 Histograma distribución triangular – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)

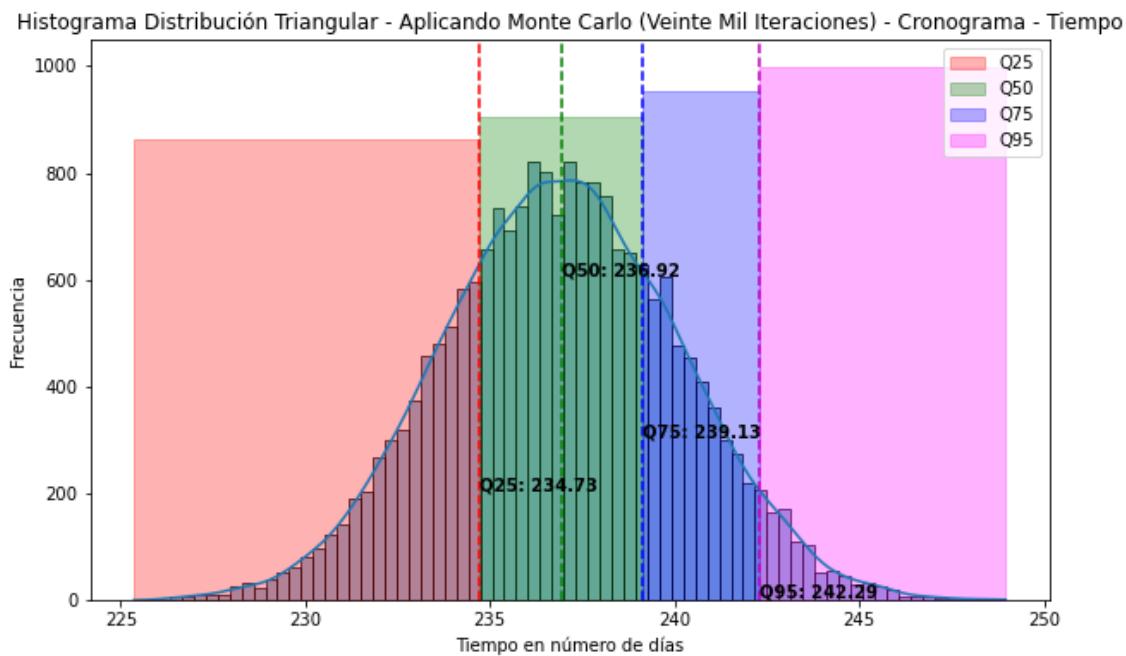
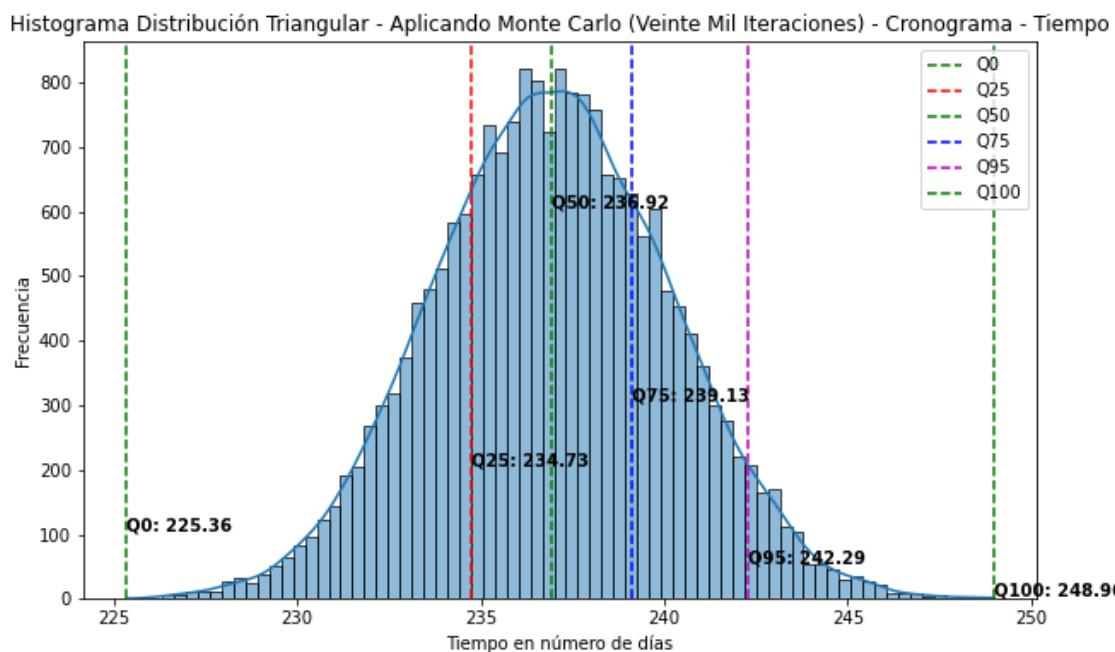


Tabla 12 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000, 10,000 y 20,000 Autor: (Osejo Francisco, 2023)

Distribución Triangular	Simulación Monte Carlo - Iteraciones				
	1.000	5.000	10.000	20.000	Media
Q0	226,884	225,758	222,516	225,360	225,130
Q25	234,720	234,790	234,712	234,73	234,738
Q50	236,840	237,010	236,866	236,923	236,910
Q75	239,050	239,200	239,084	239,130	239,116
Q95	241,960	242,330	242,273	242,292	242,214
Q100	248,371	249,386	249,582	248,960	249,075

La tabla 12 muestra la evolución obtenida en cada conjunto de iteraciones, la precisión va siendo evidente a medida que disminuye la diferencia entre los resultados obtenidos.

Gráfico 33 Histograma distribución triangular – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)



Este Gráfico muestra todos los percentiles, lo que, facilita totalmente su análisis.

Tabla 13 Cuadro comparativo de información levantada en cada iteración aplicando monte carlo conjuntamente con la distribución triangular Autor: (Osejo Francisco, 2023)

Distribución Triangular	Simulación Monte Carlo - Iteraciones				Media
	1.000	5.000	10.000	20.000	
Frecuencia	110	330	500	830	442,5
Media	236,84	237,01	236,93	236,904	236,921
Desviación estándar	3,1864	3,2535	3,333	3,262	3,259
Tiempo ejecución segundos	0,53	1,64	2,3	2,45	1,73
Reserva en número de días	1,84	2,01	1,93	1,90	1,921
Q0	226,884	225,76	222,516	225,360	225,130
Q25	234,720	234,790	234,712	234,73	234,738
Q50	236,840	237,010	236,866	236,923	236,910
Q75	239,050	239,200	239,084	239,130	239,116
Q95	241,960	242,330	242,273	242,292	242,214
Q100	248,371	249,386	249,582	248,960	249,075

Probabilidad de cumplir el cronograma objetivo, estimación Más Probable 235 días:	28%	27,24%	27,33%	27,65%	27,6%
Probabilidad de cumplir el cronograma objetivo, estimación Optimista 207,28 días:	0%	0%	0%	0%	0%
Probabilidad de cumplir el cronograma objetivo, estimación Pesimista 268,50 días:	100%	100%	100%	100%	100%
Comprobación Q0, Q25, Q50, Q75, Q95, Q100					
Probabilidad Q0:	0,00%	0,00%	0,00%	0,00%	0%
Probabilidad Q25:	25,20%	24,94%	24,99%	25%	25%
Probabilidad Q50:	50%	49,88%	50%	49,99%	50%
Probabilidad Q75:	75%	74,90%	74,99%	74,99%	75%
Probabilidad Q95:	94,80%	94,92%	94,99%	95,00%	95%
Probabilidad Q100:	99,90%	99,98%	99,99%	99,99%	100%

Conclusión:

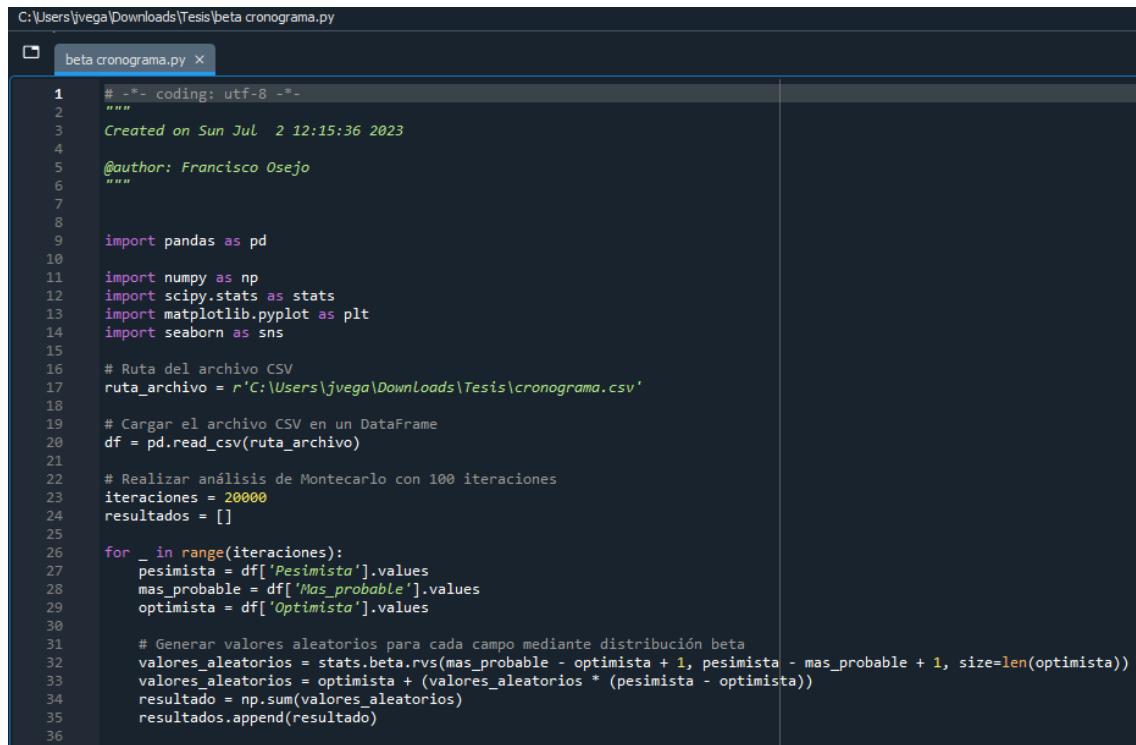
- Muestra la duración mínima en días **225,130**, así como la duración máxima en días **249,075**, a más de la duración más probable en días o Q50 **236,921**.
- Permite cuantificar el nivel de riesgo existente.
- Facilita determinar qué actividades tienen mayor riesgo
- Suministra información vital para estructurar los respectivos planes de contingencia
- Facilita visualizar las oportunidades – riesgos positivos.
- Ayuda a determinar las reservas de tiempo
- Obtener y analizar la desviación estándar, por ejemplo:
 - Media de Q50: **236,910 días (+ -) 3,259 días**
 - Rango posible entre: **233,651 días a 240,169 días**
- El valor más probable es de 235 días que equivale 27,60% de probabilidad de lograr el objetivo, lo que indica que existe alto riesgo, habría que revisar las estimaciones para ver dónde está el error y hacer las correcciones del caso.

- El valor optimista es de 207,28 días que equivale al 0% de probabilidad de lograr el objetivo, lo que indica que existe alto riesgo, por lo tanto, no se lo debería tomar en cuenta.
- El valor pesimista 268,50 días equivale al 100% de probabilidad de lograr el objetivo, pero no es determinante, ya que, la media de Q100 es 249 días, por lo tanto, valor pesimista tampoco es buen medidor, lo podríamos considerar un sobreajuste del modelo.
- Por lo tanto, la estimación de la duración en días hechas en el análisis inicial Vs el resultado de la simulación de Monte Carlo permite determinar que lo estimado inicialmente no es confiable, y se debe volver a reformular en base al resultado obtenido con Monte Carlo.
- Las estimaciones de la duración original Vs la simulación de Montecarlo obtenida determinan que existe un riesgo muy alto de lograr el objetivo, por lo tanto, es recomendable volver revisar los riesgos y reformular las estimaciones de la duración para lograr encontrar el modelo adecuado.
- Es necesario reformular las estimaciones de tiempo y volver a iterar con Monte Carlo para encontrar el modelo adecuado.

3.5.2.1.1.2 Distribución Probabilística Beta o Pert

3.5.2.1.1.2.1 Código en Python Distribución Probabilística Beta o Pert

Gráfico 34 Simulación monte carlo - código en Python distribución beta o pert cronograma – tiempo Autor: (Osejo Francisco, 2023)



The screenshot shows a code editor window with the file 'beta cronograma.py' open. The code is written in Python and performs a Monte Carlo simulation using the Beta distribution to generate values between the 'Mas_probable' and 'Optimista' bounds defined in the 'cronograma.csv' file. The code includes imports for pandas, numpy, scipy.stats, matplotlib.pyplot, and seaborn. It reads a CSV file, sets up 2000 iterations, and uses a for loop to generate random beta-distributed values for each iteration, summing them up to get the final result.

```
# -*- coding: utf-8 -*-
"""
Created on Sun Jul  2 12:15:36 2023
@author: Francisco Osejo
"""

import pandas as pd
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns

# Ruta del archivo CSV
ruta_archivo = r'C:\Users\jvega\Downloads\Tesis\cronograma.csv'

# Cargar el archivo CSV en un DataFrame
df = pd.read_csv(ruta_archivo)

# Realizar análisis de Montecarlo con 100 iteraciones
iteraciones = 2000
resultados = []

for _ in range(iteraciones):
    pesimista = df['Pesimista'].values
    mas_probable = df['Mas_probable'].values
    optimista = df['Optimista'].values

    # Generar valores aleatorios para cada campo mediante distribución beta
    valores_aleatorios = stats.beta.rvs(mas_probable - optimista + 1, pesimista - mas_probable + 1, size=len(optimista))
    valores_aleatorios = optimista + (valores_aleatorios * (pesimista - optimista))
    resultado = np.sum(valores_aleatorios)
    resultados.append(resultado)
```

Gráfico 35 Simulación monte carlo - código en Python distribución beta o pert cronograma – tiempo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\beta cronograma.py
beta cronograma.py ×

37
38     # Calcular los percentiles
39     q0 = np.percentile(resultados, 0)
40     q25 = np.percentile(resultados, 25)
41     q50 = np.percentile(resultados, 50)
42     q75 = np.percentile(resultados, 75)
43     q95 = np.percentile(resultados, 95)
44     q100 = np.percentile(resultados, 100)
45
46     # Imprimir los resultados
47     print("Q0:", q0)
48     print("Q25:", q25)
49     print("Q50 (Mediana):", q50)
50     print("Q75:", q75)
51     print("Q95:", q95)
52     print("Q100:", q100)
53
54     # Crear histograma
55     plt.figure(figsize=(10, 6))
56     sns.histplot(resultados, kde=True)
57
58     # Agregar líneas verticales en los percentiles
59     plt.axvline(q0, color='g', linestyle='--', label='Q0')
60     plt.axvline(q25, color='r', linestyle='--', label='Q25')
61     plt.axvline(q50, color='g', linestyle='--', label='Q50')
62     plt.axvline(q75, color='b', linestyle='--', label='Q75')
63     plt.axvline(q95, color='m', linestyle='--', label='Q95')
64     plt.axvline(q100, color='g', linestyle='--', label='Q100')
65
66     # Agregar anotaciones con los valores de los percentiles en negrita y color negro
67     plt.text(q0, 100, f'Q0: {q0:.2f}', rotation=0, va='bottom', color='black', weight='bold')
68     plt.text(q25, 200, f'Q25: {q25:.2f}', rotation=0, va='bottom', color='black', weight='bold')
69     plt.text(q50, 500, f'Q50: {q50:.2f}', rotation=0, va='bottom', color='black', weight='bold')
70     plt.text(q75, 300, f'Q75: {q75:.2f}', rotation=0, va='bottom', color='black', weight='bold')
71     plt.text(q95, 50, f'Q95: {q95:.2f}', rotation=0, va='bottom', color='black', weight='bold')
72     plt.text(q100, 0, f'Q0: {q100:.2f}', rotation=0, va='bottom', color='black', weight='bold')
73
```

Gráfico 36 Simulación monte carlo - código en Python distribución beta o pert cronograma – tiempo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\beta cronograma.py
beta cronograma.py

74 plt.xlabel('Tiempo en número de días')
75 plt.ylabel('Frecuencia')
76 plt.title('Histograma Distribución Beta o Pert - Aplicando Monte Carlo (Veinte Mil Iteraciones) - Cronograma - Tiempo')
77 plt.legend()
78 plt.show()
79
80 # Crear histograma mostrando divisiones de nivel de riesgo
81 plt.figure(figsize=(10, 6))
82 sns.histplot(resultados, kde=True)
83
84 # Rellenar áreas con diferentes colores
85 plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q0, q25, alpha=0.3, color='red', label='Q25')
86 plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q25, q75, alpha=0.3, color='green', label='Q50')
87 plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q75, q95, alpha=0.3, color='blue', label='Q75')
88 plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q95, q100, alpha=0.3, color='magenta', label='Q95')
89
90 # Agregar líneas verticales en los percentiles
91 plt.axvline(q25, color='r', linestyle='--')
92 plt.axvline(q50, color='g', linestyle='--')
93 plt.axvline(q75, color='b', linestyle='--')
94 plt.axvline(q95, color='m', linestyle='--')
95
96 # Agregar anotaciones con los valores de los percentiles en negrita y color negro
97 plt.text(q25, 200, f'Q25: {q25:.2f}', rotation=0, va='bottom', color='black', weight='bold')
98 plt.text(q50, 500, f'Q50: {q50:.2f}', rotation=0, va='bottom', color='black', weight='bold')
99 plt.text(q75, 300, f'Q75: {q75:.2f}', rotation=0, va='bottom', color='black', weight='bold')
100 plt.text(q95, 0, f'Q95: {q95:.2f}', rotation=0, va='bottom', color='black', weight='bold')
101
102 plt.xlabel('Tiempo en número de días')
103 plt.ylabel('Frecuencia')
104 plt.title('Histograma Distribución Beta o Pert - Aplicando Monte Carlo (Veinte Mil Iteraciones) - Cronograma - Tiempo')
105 plt.legend()
106 plt.show()
107
108
```

Gráfico 37 Simulación monte carlo - código en Python distribución beta o pert cronograma – tiempo Autor: (Osejo Francisco, 2023)

```
108 # Calcular la media y la desviación estándar de los resultados
109 media = np.mean(resultados)
110 desviacion_estandar = np.std(resultados)
111
112 # Calcular la probabilidad de cumplir el presupuesto objetivo
113 # Presupuesto objetivo
114 presupuesto_objetivo = 240.818
115 probabilidad_cumplir = sum(resultado <= presupuesto_objetivo for resultado in resultados) / iteraciones
116
117 print("Probabilidad de cumplir el presupuesto objetivo:", probabilidad_cumplir)
118
```

Se realizó la simulación de Monte Carlo para 1.000, 5.000, 10.000 y 20.000 iteraciones aplicando stats.beta.rvs () de scipy de manera independiente.

```
# Generar valores aleatorios para cada campo mediante distribución beta

valores_aleatorios = stats.beta.rvs(mas_probable - optimista + 1, pesimista -
mas_probable + 1, size=len(optimista))

valores_aleatorios = optimista + (valores_aleatorios * (pesimista - optimista))

resultado = np.sum(valores_aleatorios)
```

resultados.append(resultado) (Osejo Francisco, 2023).

Para efectos de este trabajo de titulación sólo mostraremos el código aplicado a las 20,000 iteraciones.

Para las 1.000, 5.000, 10.000 iteraciones hay que cambiar en el código el valor de la variable “iteraciones =”, adicionalmente cambiar posiciones y textos en el código correspondiente a la generación de los histogramas respectivos en cada ejecución del programa. (Osejo Francisco, 2023).

3.5.2.1.1.2.2 Histogramas obtenidos en Python Distribución Probabilística Beta o Pert

Para efectos de este trabajo de titulación solo mostraremos Histogramas obtenidos en las simulaciones de 10,000 y 20,000 iteraciones

Gráfico 38 Histograma distribución beta o pert – aplicando monte carlo (diez mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)

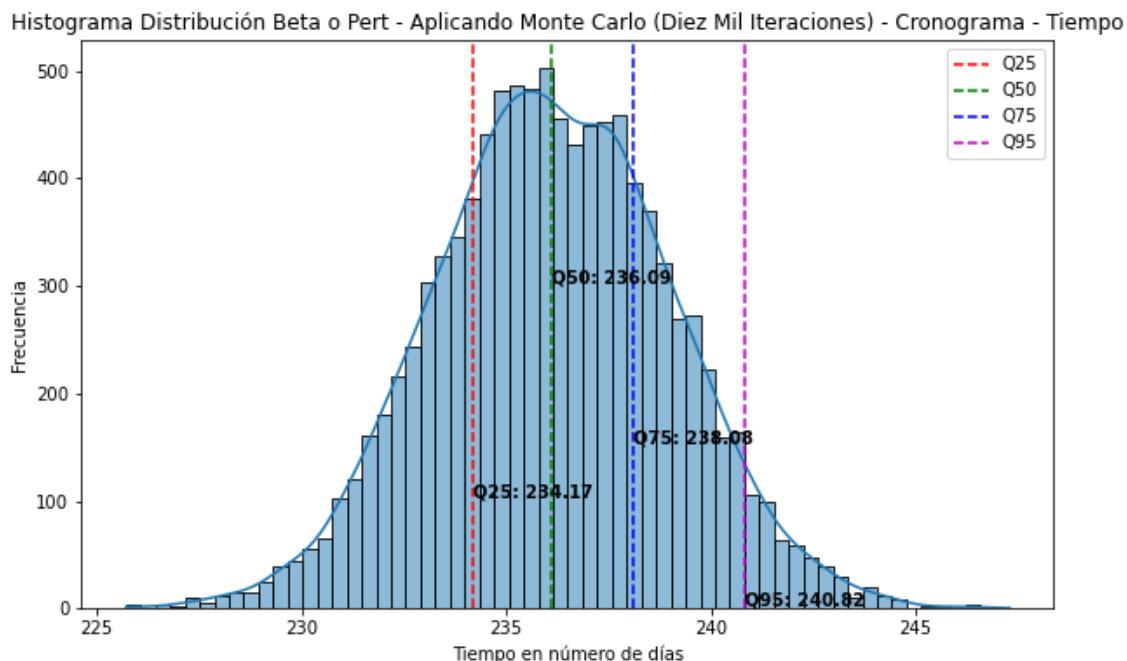
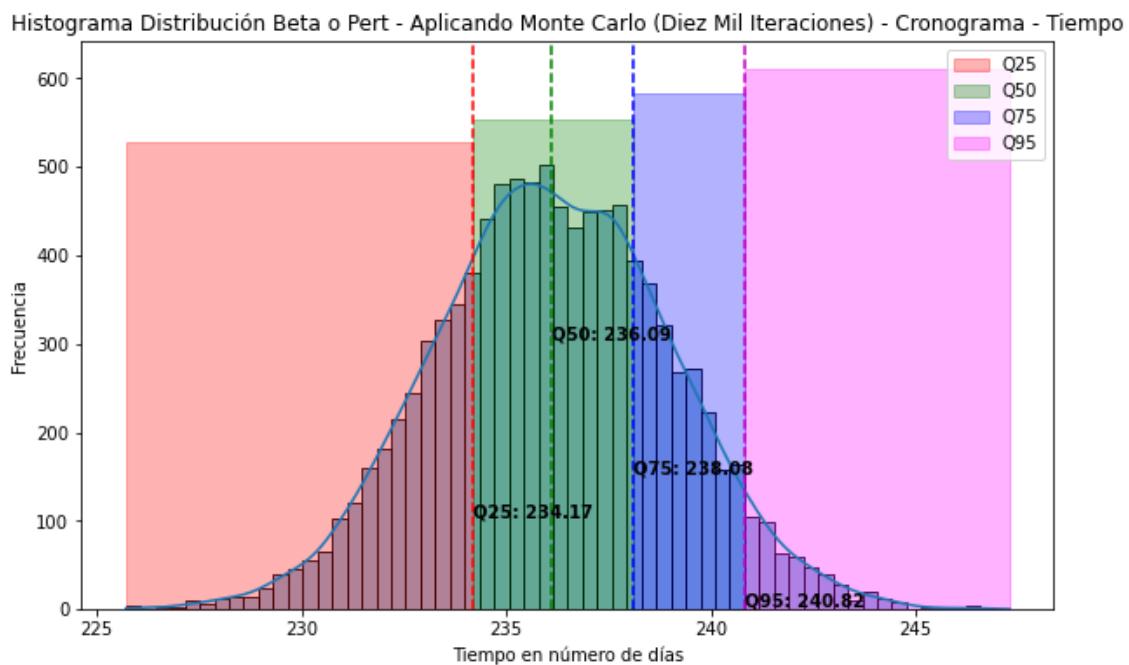


Gráfico 39 Histograma distribución beta o pert – aplicando monte carlo (diez mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)



Como podemos observar en los Gráficos 38 y 39 de este documento, el código del algoritmo realiza 10,000 diez mil iteraciones en donde aplicamos stats.beta.rvs () de la librería scipy.

```
valores_aleatorios = stats.beta.rvs(mas_probable - optimista + 1, pesimista - mas_probable + 1, size=len(optimista)) (Osejo Francisco, 2023)
```

Realiza los cálculos en base los valores generados randomicamente posibles dentro de los rangos de los valores de las variables optimista y pesimista (Osejo Francisco, 2023).

Obtenemos los percentiles Q0, Q25, Q50, Q75, Q95, Q100:

Tabla 14 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000 y 10,000 Autor: (Osejo Francisco, 2023)

Distribución Beta-Pert	Simulación Monte Carlo - Iteraciones		
	1.000	5.000	10.000
Q0	227,245	225,852	225,714
Q25	234,117	234,228	234,173
Q50	236,207	236,064	236,092
Q75	238,126	238,008	238,083

Q95	240,931	240,87	240,823
Q100	246,532	246,952	247,290

Q0: Muestra el valor mínimo en días que pueden iniciar las actividades

Q50: Muestra el valor medio en días en los cuales las actividades pueden ejecutarse

Q100: Muestra el valor máximo en días que pueden terminar las actividades

La interpretación de los Gráficos 38 y 39 es la siguiente:

- Permite identificar visualmente los límites de los rangos de tiempos en días
- Facilita el análisis de estimaciones de tiempo en días
- Muestra con mayor precisión las probabilidades de lograr el objetivo
 - Q0: Cualquier duración menor o igual al Q0 tendrá 0% de probabilidades de lograr el objetivo.
 - Q25: Cualquier duración menor o igual al Q25 tendrá de 0% a 25% de probabilidades de lograr el objetivo.
 - Q50: Cualquier duración menor o igual al Q50 tendrá de 26% a 50% de probabilidades de lograr el objetivo.
 - Q75: Cualquier duración menor o igual al Q75 tendrá de 51% a 75% de probabilidades de lograr el objetivo.
 - Q100: Cualquier duración menor o igual al Q100 tendrá de 76% a 100% de probabilidades de lograr el objetivo.
- Obtener los percentiles nos permite determinar el nivel de riesgo existente en el cumplimiento de los objetivos.
 - Q25: Indica que existe un 25% de probabilidad de éxito, lo que implica un alto riesgo de lograr el objetivo.
 - Q50: Indica que existe un 50% de probabilidad de éxito, lo que implica un riesgo medio de lograr el objetivo.
 - Q75: Indica que existe un 75% de probabilidad de éxito, lo que implica un riesgo bajo de lograr el objetivo.
 - Q100: Indica que existe un 100% de probabilidad de éxito, lo que implica que no existe riesgo de lograr el objetivo.

- Conseguir las ponderaciones de riesgo (Alto, medio, bajo) conjuntamente con las probabilidades de éxito permite determinar la matriz de probabilidad e impacto con esquema de puntuación que se muestra en el Gráfico 5
- La Frecuencia muestra claramente el nivel de iteraciones que coincidieron en ese rango de análisis. Por lo que, es muy importante revisar los resultados obtenidos en cada conjunto de iteraciones.
- La información descrita en la Tabla 14 muestra que a mayor número de iteraciones (1000, 5000, 10000) se logra afinar la precisión en los resultados obtenidos.

Gráfico 40 Histograma distribución beta o pert – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)

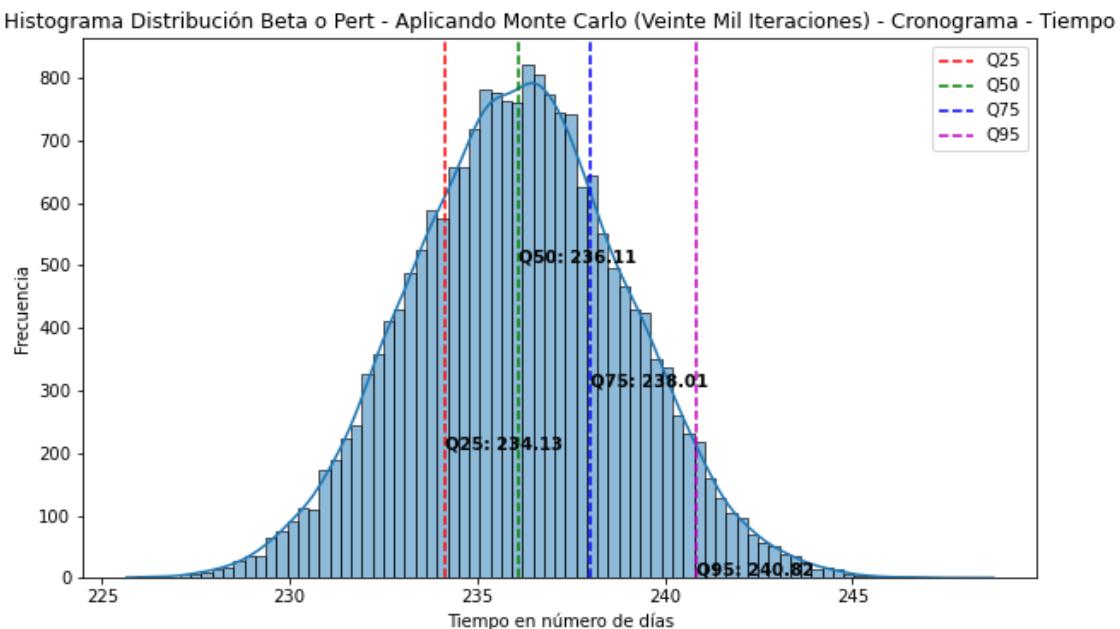


Gráfico 41 Histograma distribución beta o pert – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)

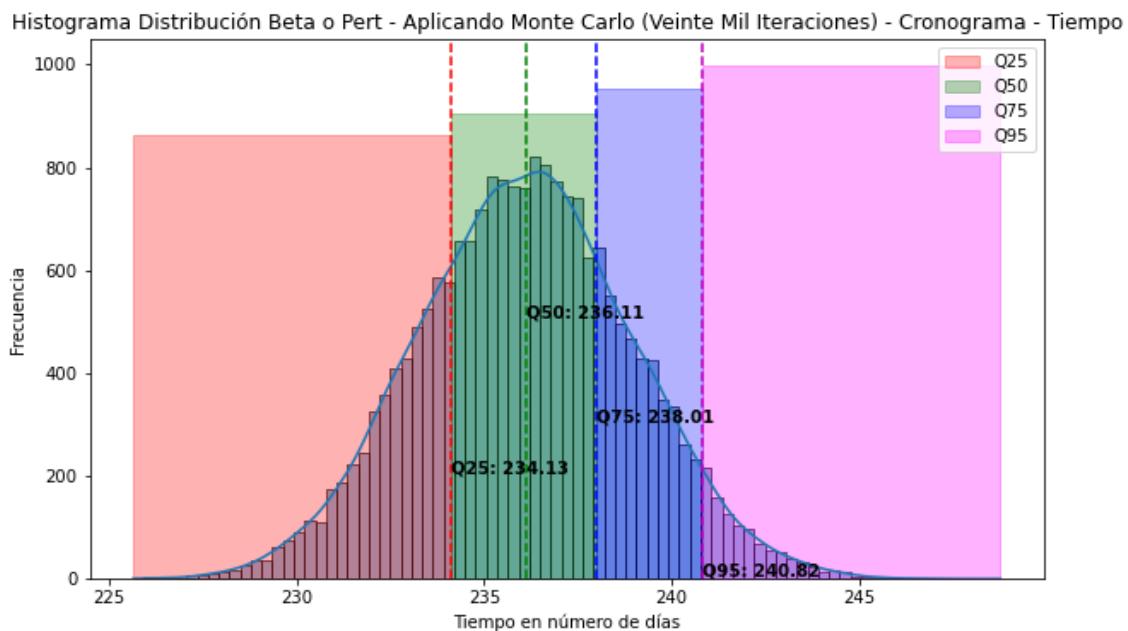
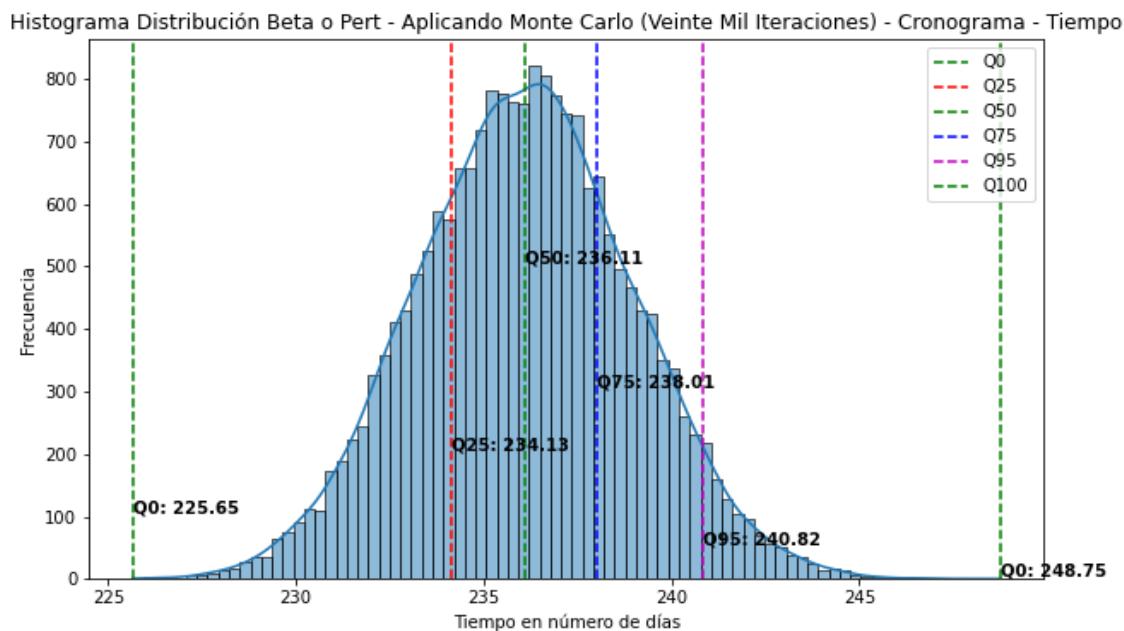


Tabla 15 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000, 10,000 y 20,000 Autor: (Osejo Francisco, 2023)

Distribución Beta-Pert	Simulación Monte Carlo - Iteraciones				
	1.000	5.000	10.000	20.000	Media
Q0	227,245	225,852	225,714	225,646	226,114
Q25	234,117	234,228	234,173	234,134	234,163
Q50	236,207	236,064	236,092	236,107	236,118
Q75	238,126	238,008	238,083	238,010	238,057
Q95	240,931	240,87	240,823	240,818	240,861
Q100	246,532	246,952	247,290	248,753	247,382

La tabla 15 muestra la evolución obtenida en cada conjunto de iteraciones, la precisión va siendo evidente a medida que disminuye la diferencia entre los resultados obtenidos.

Gráfico 42 Histograma distribución beta o pert – aplicando monte carlo (veinte mil iteraciones) – cronograma – tiempo Autor: (Osejo Francisco, 2023)



Este Gráfico muestra todos los percentiles, lo que, facilita totalmente su análisis.

Tabla 16 Cuadro comparativo de información levantada en cada iteración aplicando monte carlo conjuntamente con la distribución beta o pert Autor: (Osejo Francisco, 2023)

Distribución Beta-Pert	Simulación Monte Carlo - Iteraciones				Media
	1.000	5.000	10.000	20.000	
Frecuencia	120	320	500	800	435
Media	236,187	236,132	236,124	236,092	236,134
Desviación estándar	2,872	2,849	2,90	2,87	2,872
Tiempo ejecución segundos	1,15	1,96	4,94	5,22	3,318
Reserva en número de días	1,19	1,13	1,12	1,09	1,134
Q0	227,245	225,852	225,714	225,646	226,114
Q25	234,117	234,228	234,173	234,134	234,163
Q50	236,207	236,064	236,092	236,107	236,118
Q75	238,126	238,008	238,083	238,010	238,057
Q95	240,931	240,87	240,823	240,818	240,861
Q100	246,532	246,952	247,290	248,753	247,382
Probabilidad de cumplir el cronograma objetivo, estimación Más Probable 235 días :	33,40%	35,08%	35,21%	35,17%	34,72%

Probabilidad de cumplir el cronograma objetivo, estimación Optimista 207,28 días :	0%	0%	0%	0%	0%
Probabilidad de cumplir el cronograma objetivo, estimación Pesimista 268,50 días :	100%	100%	100%	100%	100%
Comprobación Q0,Q25, Q50, Q75, Q95,Q100					
Probabilidad Q0:	0,00%	0,00%	0,00%	0,00%	0,00%
Probabilidad Q25:	25%	25%	24,99%	24,99%	25,00%
Probabilidad Q50:	49,99%	50%	49,99%	49,99%	49,99%
Probabilidad Q75:	75%	74,98%	74,94%	74,99%	74,98%
Probabilidad Q95:	94,90%	94,98%	94,99%	94,99%	94,97%
Probabilidad Q100:	99,99%	99,98%	99,99%	99,99%	99,99%

Conclusión:

- Muestra la duración mínima en días **226,114**, así como la duración máxima en días **247,382**, a más de la duración más probable en días o **Q50 236,134**.
- Permite cuantificar el nivel de riesgo existente.
- Facilita determinar qué actividades tienen mayor riesgo
- Suministra información vital para estructurar los respectivos planes de contingencia
- Facilita visualizar las oportunidades – riesgos positivos.
- Ayuda a determinar las reservas de tiempo
- Obtener y analizar la desviación estándar, por ejemplo:
 - Media de Q50: **236,134 días (+ -) 2,872 días**
 - Rango posible entre: **233,262 días a 239,006 días**
- El valor más probable es de 235 días que equivale 34,72% de probabilidad de lograr el objetivo, lo que indica que existe alto riesgo, habría que revisar las estimaciones para ver dónde está el error y hacer las correcciones del caso.
- El valor optimista es de 207,28 días que equivale al 0% de probabilidad de lograr el objetivo, lo que indica que existe alto riesgo, por lo tanto, no se lo debería tomar en cuenta.
- El valor pesimista 268,50 días equivale al 100% de probabilidad de lograr el objetivo, pero no es determinante, ya que, la media de Q100 es **247** días, por lo tanto, valor

pesimista tampoco es buen medidor, lo podríamos considerar un sobreajuste del modelo.

- Por lo tanto, la estimación de la duración en días hechas en el análisis inicial Vs el resultado de la simulación de Monte Carlo permite determinar que lo estimado inicialmente no es confiable, y se debe volver a reformular en base al resultado obtenido con Monte Carlo.
- Las estimaciones de la duración original Vs la simulación de Montecarlo obtenida determinan que existe un riesgo muy alto de lograr el objetivo, por lo tanto, es recomendable volver revisar los riesgos y reformular las estimaciones de la duración para lograr encontrar el modelo adecuado.
- Es necesario reformular las estimaciones de tiempo y volver a iterar con Monte Carlo para encontrar el modelo adecuado.

3.5.2.1.2 Aplicar Monte Carlo a Dataset de Presupuesto – Costo

Gráfico 43 Simulación monte carlo – contenido dataset presupuesto – costo Autor: (Osejo Francisco, 2023)

Index	Descripción	Tiempo	Otimista	as probal	Pesimista	ucion Tri	ribucion l
0	FIDS/BIDS/RMS/AODB/CUTE - Lumpsum (Ultra) ->	292.88	982070	984171	986272	984171	984171
1	Servers y SO->	27	78549	92410.6	104424	91794.5	92102.6
2	Workstation (HP) ->	45	78213.8	85015	96917.1	86715.3	85865.1
3	DDC (ACER) ->	72	73320	78000	92040	81120	79560
4	Monitores 42->	71	96360	109500	122640	109500	109500
5	Brackets for FIDS Monitors + Imstallation ->	65	44500	50000	58000	50833.3	50416.7
6	LAN and VoIP - Materiales (Cisco / DOS) ->	194	974763	975713	983734	978070	976892
7	LAN and VoIP - Servicios (Cisco / DOS) ->	444.63	190144	208950	231934	210342	209646
8	Master Clock (Moba Time) ->	230	9310.95	10120.6	11638.7	10356.7	10238.7
9	FIDS Labor JCI ->	188.88	323921	344597	406625	358381	351489
10	FIDS FMT JCI ->	64	286668	329503	382223	332798	331151
11	Security Systems: Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	287	975718	983394	995238	984783	984088
12	Subcontract Installation ->	89	66750	75000	85500	75750	75375
13	Security Labor JCI ->	104	190000	200000	226000	205333	202667
14	Security FMT JCI ->	50	106814	113632	123859	114768	114200
15	Public Address (Bosh) ->	90	353400	380000	399000	377467	378733
16	Ingenieria PA (Bosh) ->	30	31850	35000	39200	35350	35175
17	Notifier ->	120	246298	256560	289913	264257	260408
18	Wattstopper ->	180	37600	40000	45600	41066.7	40533.3
19	PA Labor JCI ->	110	105411	114578	127181	115723	115150
20	PA FMT JCI ->	90	72428.7	79592	94714.5	82245.1	80918.5

3.5.2.1.2.1 Distribución Probabilística Triangular

3.5.2.1.2.1.1 Código en Python Distribución Probabilística Triangular

Gráfico 44 Simulación monte carlo - código en Python distribución probabilística triangular presupuesto – costo Autor: (Osejo Francisco, 2023)

The screenshot shows a code editor window with the file name 'Triangular Costo.py' at the top. The code itself is a Python script for performing a Monte Carlo simulation. It starts by importing pandas, numpy, matplotlib.pyplot, and seaborn. It then defines the path to a CSV file ('Costos.csv') and reads it into a DataFrame. The script then performs 2000 iterations, generating random numbers from a triangular distribution for each iteration based on minimum, most probable, and maximum values from the CSV. Finally, it calculates the total sum of these random values for each iteration and appends it to a list of results.

```
C:\Users\jvega\Downloads\Tesis\Triangular Costo.py
  □ Triangular Costo.py* ×
1  # -*- coding: utf-8 -*-
2  """
3  Created on Sun Jul  2 17:36:48 2023
4
5  @author: Francisco Osejo
6  """
7
8  import pandas as pd
9  import numpy as np
10 import matplotlib.pyplot as plt
11 import seaborn as sns
12
13 # Ruta del archivo CSV
14
15 ruta_archivo = r'C:\Users\jvega\Downloads\Tesis\Costos.csv'
16
17 # Cargar el archivo CSV en un DataFrame
18
19 df = pd.read_csv(ruta_archivo, sep=';', decimal=',')
20 #df = pd.read_csv(ruta_archivo)
21
22 # Realizar análisis de Montecarlo con 100 iteraciones
23 iteraciones = 20000
24 resultados = []
25
26
27 for _ in range(iteraciones):
28     pesimista = df['Pesimista'].values
29     mas_probable = df['Mas_probable'].values
30     optimista = df['Optimista'].values
31
32     # Generar valores aleatorios para cada campo mediante distribución triangular
33     valores_aleatorios = np.random.triangular(optimista,mas_probable,pesimista)
34
35     # Calcular el resultado final de la iteración
36     resultado = np.sum(valores_aleatorios)
37     resultados.append(resultado)
```

Gráfico 45 Simulación monte carlo - código en Python distribución probabilística triangular presupuesto – costo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\Triangular Costo.py
  □ Triangular Costo.py* ×
38
39     # Calcular los percentiles
40     q0 = np.percentile(resultados, 0)
41     q25 = np.percentile(resultados, 25)
42     q50 = np.percentile(resultados, 50)
43     q75 = np.percentile(resultados, 75)
44     q95 = np.percentile(resultados, 95)
45     q100 = np.percentile(resultados, 100)
46
47     # Imprimir los resultados
48     print("Q0:", q0)
49     print("Q25:", q25)
50     print("Q50 (Mediana):", q50)
51     print("Q75:", q75)
52     print("Q95:", q95)
53     print("Q100:", q100)
54
55     # Crear histograma
56     plt.figure(figsize=(10, 6))
57     sns.histplot(resultados, kde=True)
58
59     # Agregar líneas verticales en los percentiles
60     plt.axvline(q0, color='g', linestyle='--', label='Q0')
61     plt.axvline(q25, color='r', linestyle='--', label='Q25')
62     plt.axvline(q50, color='g', linestyle='--', label='Q50')
63     plt.axvline(q75, color='b', linestyle='--', label='Q75')
64     plt.axvline(q95, color='m', linestyle='--', label='Q95')
65     plt.axvline(q100, color='g', linestyle='--', label='Q100')
66
67     # Agregar anotaciones con los valores de los percentiles en negrita y color negro
68     plt.text(q0, 0, f'Q0: {q0:.2f}', rotation=0, va='bottom', color='black', weight='bold')
69     plt.text(q25, 100, f'Q25: {q25:.2f}', rotation=0, va='bottom', color='black', weight='bold')
70     plt.text(q50, 500, f'Q50: {q50:.2f}', rotation=0, va='bottom', color='black', weight='bold')
71     plt.text(q75, 300, f'Q75: {q75:.2f}', rotation=0, va='bottom', color='black', weight='bold')
72     plt.text(q95, 100, f'Q95: {q95:.2f}', rotation=0, va='bottom', color='black', weight='bold')
73     plt.text(q100, 0, f'Q100: {q100:.2f}', rotation=0, va='bottom', color='black', weight='bold')
```

Gráfico 46 Simulación monte carlo - código en Python distribución probabilística triangular presupuesto – costo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\Triangular Costo.py
  □ Triangular Costo.py* ×
74     plt.xlabel('Presupuesto en millones de dólares americanos')
75     plt.ylabel('Frecuencia')
76     plt.title('Histograma Distribución Triangular - Aplicando Monte Carlo (Veinte Mil Iteraciones) Presupuesto - Costo')
77     plt.legend()
78     plt.show()
79
80
81     # Graficar el histograma
82     plt.figure(figsize=(10, 6))
83     sns.histplot(resultados, kde=True)
84
85     # Rellenar áreas con diferentes colores
86     plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q0, q25, alpha=0.3, color='red', label='Q25')
87     plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q25, q75, alpha=0.3, color='green', label='Q50')
88     plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q75, q95, alpha=0.3, color='blue', label='Q75')
89     plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q95, q100, alpha=0.3, color='magenta', label='Q95')
90
91     # Agregar líneas verticales en los percentiles
92     plt.axvline(q25, color='r', linestyle='--')
93     plt.axvline(q50, color='g', linestyle='--')
94     plt.axvline(q75, color='b', linestyle='--')
95     plt.axvline(q95, color='m', linestyle='--')
96
97     # Agregar anotaciones con los valores de los percentiles en negrita y color negro
98     plt.text(q25, 100, f'Q25: {q25:.2f}', rotation=0, va='bottom', color='black', weight='bold')
99     plt.text(q50, 500, f'Q50: {q50:.2f}', rotation=0, va='bottom', color='black', weight='bold')
100    plt.text(q75, 300, f'Q75: {q75:.2f}', rotation=0, va='bottom', color='black', weight='bold')
101    plt.text(q95, 0, f'Q95: {q95:.2f}', rotation=0, va='bottom', color='black', weight='bold')
102
103    plt.xlabel('Presupuesto en millones de dólares americanos')
104    plt.ylabel('Frecuencia')
105    plt.title('Histograma Distribución Triangular - Aplicando Monte Carlo (Veinte Mil Iteraciones) Presupuesto - Costo')
106    plt.legend()
107    plt.show()
```

Gráfico 47 Simulación monte carlo - código en Python distribución probabilística triangular presupuesto – costo Autor: (Osejo Francisco, 2023)

```
108  
109  
110     # Calcular la media y la desviación estándar de los resultados  
111     media = np.mean(resultados)  
112     desviacion_estandar = np.std(resultados)  
113  
114     # Calcular la probabilidad de cumplir el presupuesto objetivo  
115     # Presupuesto objetivo  
116     presupuesto_objetivo = 5545735.24  
117     probabilidad_cumplir = sum(resultado <= presupuesto_objetivo for resultado in resultados) / iteraciones  
118  
119     print("Probabilidad de cumplir el presupuesto objetivo:", probabilidad_cumplir)  
120
```

Se realizó la simulación de Monte Carlo para 1.000, 5.000, 10.000 y 20.000 iteraciones aplicando `np.random.triangular()` de numpy de manera independiente.

```
# Generar valores aleatorios para cada campo mediante distribución triangular  
  
valores_aleatorios = np.random.triangular(optimista,mas_probable,pesimista) (Osejo  
Francisco, 2023)
```

Para efectos de este trabajo de titulación sólo mostraremos el código aplicado a las 20,000 iteraciones.

Para las 1.000, 5.000, 10.000 iteraciones hay que cambiar en el código el valor de la variable “iteraciones =”, adicionalmente cambiar posiciones y textos en el código correspondiente a la generación de los histogramas respectivos en cada ejecución del programa. (Osejo Francisco, 2023)

3.5.2.1.2.1.2 Histogramas obtenidos en Python Distribución Probabilística Triangular

Para efectos de este trabajo de titulación solo mostraremos Histogramas obtenidos en las simulaciones de 10,000 y 20,000 iteraciones

Gráfico 48 Histograma Distribución Triangular – Aplicando Monte Carlo (Diez Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)

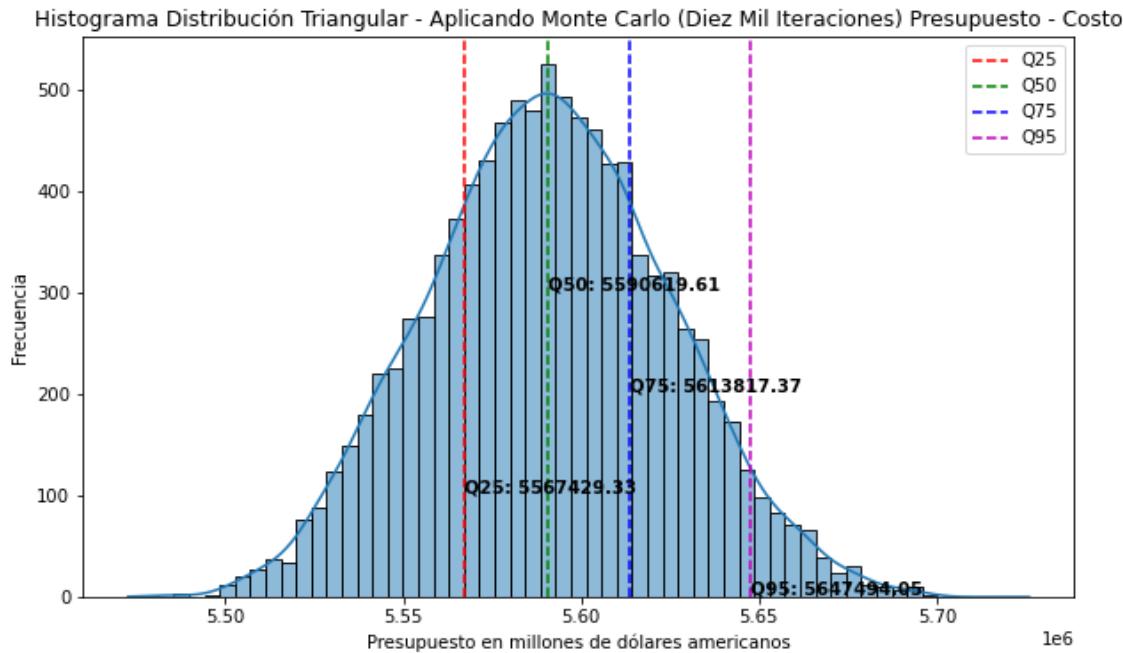
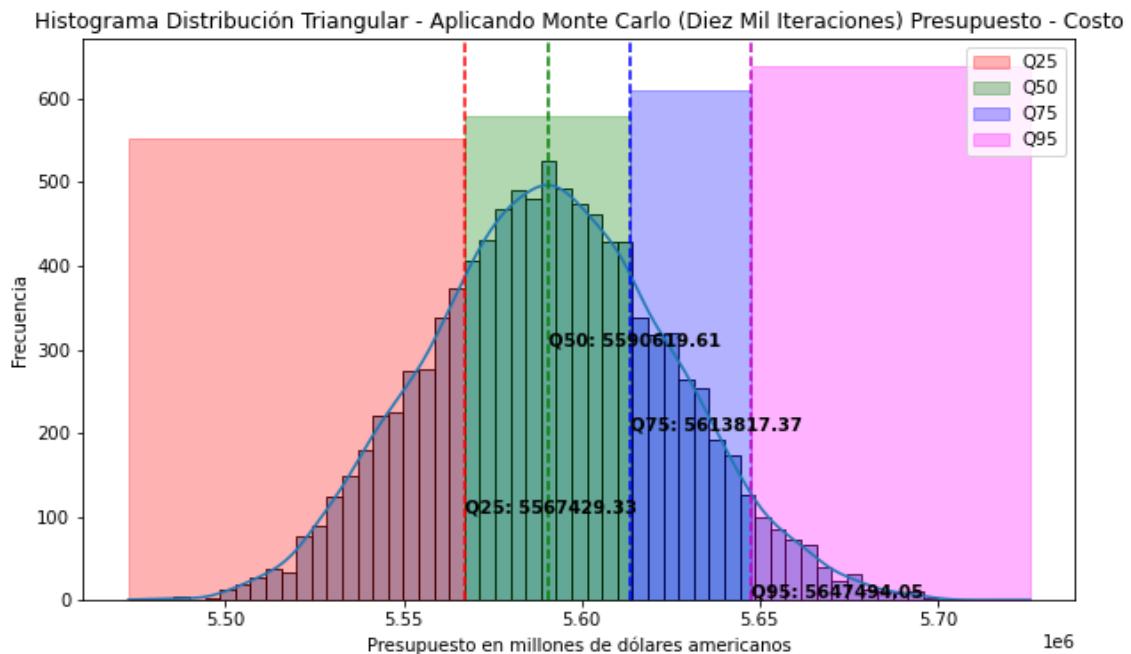


Gráfico 49 Histograma Distribución Triangular – Aplicando Monte Carlo (Diez Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)



Como podemos observar en los Gráficos 48 y 49 de este documento, el código del algoritmo realiza 10,000 diez mil iteraciones en donde aplicamos la fórmula de la

Distribución triangular ($cE = (cO + cM + cP) / 3$), obteniendo la media de la sumatoria de los valores (optimista + más probable + pesimista) / 3, la librería de numpy np.random.triangular(optimista,mas_probable,pesimista) realiza los cálculos en base los valores generados randomicamente posibles dentro de los rangos de los valores de las variables optimista y pesimista (Osejo Francisco, 2023).

Obtenemos los percentiles Q0, Q25, Q50, Q75, Q95, Q100:

Tabla 17 Costo en millones de dólares de norte américa obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000 y 10,000 Autor: (Osejo Francisco, 2023)

Distribución Triangular	Simulación Monte Carlo - Iteraciones		
	1000	5000	10000
Q0	5.480.271,26	5.473.816,12	5.472.634,51
Q25	5.565.224,64	5.566.873,86	5.567.429,33
Q50	5.590.160,35	5.589.848,50	5.590.619,61
Q75	5.614.345,36	5.612.992,37	5.613.817,37
Q95	5.648.581,97	5.647.432,69	5.647.494,05
Q100	5.716.498,66	5.715.259,08	5.725.985,25

Q0: Muestra el presupuesto mínimo en millones de dólares que pueden costar las actividades

Q50: Muestra el presupuesto medio en millones de dólares en los cuales las actividades pueden ejecutarse

Q100: Muestra el presupuesto máximo en millones de dólares que pueden terminar las actividades

La interpretación de los Gráficos 48 y 49 es la siguiente:

- Permite identificar visualmente los límites de los rangos del presupuesto en millones de dólares
- Facilita el análisis de estimaciones del presupuesto en millones de dólares
- Muestra con mayor precisión las probabilidades de lograr el objetivo

- Q0: Cualquier presupuesto menor o igual al Q0 tendrá 0% de probabilidades de lograr el objetivo.
 - Q25: Cualquier presupuesto menor o igual al Q25 tendrá de 0% a 25% de probabilidades de lograr el objetivo.
 - Q50: Cualquier presupuesto menor o igual al Q50 tendrá de 26% a 50% de probabilidades de lograr el objetivo.
 - Q75: Cualquier presupuesto menor o igual al Q75 tendrá de 51% a 75% de probabilidades de lograr el objetivo.
 - Q100: Cualquier presupuesto menor o igual al Q100 tendrá de 76% a 100% de probabilidades de lograr el objetivo.
- Obtener los percentiles nos permite determinar el nivel de riesgo existente en el cumplimiento de los objetivos.
 - Q25: Indica que existe un 25% de probabilidad de éxito, lo que implica un alto riesgo de lograr el objetivo.
 - Q50: Indica que existe un 50% de probabilidad de éxito, lo que implica un riesgo medio de lograr el objetivo.
 - Q75: Indica que existe un 75% de probabilidad de éxito, lo que implica un riesgo bajo de lograr el objetivo.
 - Q100: Indica que existe un 100% de probabilidad de éxito, lo que implica que no existe riesgo de lograr el objetivo.
- Conseguir las ponderaciones de riesgo (Alto, medio, bajo) conjuntamente con las probabilidades de éxito permite determinar la matriz de probabilidad e impacto con esquema de puntuación que se muestra en el Gráfico 6
- La Frecuencia muestra claramente el nivel de iteraciones que coincidieron en ese rango de análisis. Por lo que, es muy importante revisar los resultados obtenidos en cada conjunto de iteraciones.
- La información descrita en la Tabla 17 muestra que a mayor número de iteraciones (1000, 5000, 10000) se logra afinar la precisión en los resultados obtenidos.

Gráfico 50 Histograma Distribución Triangular – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)

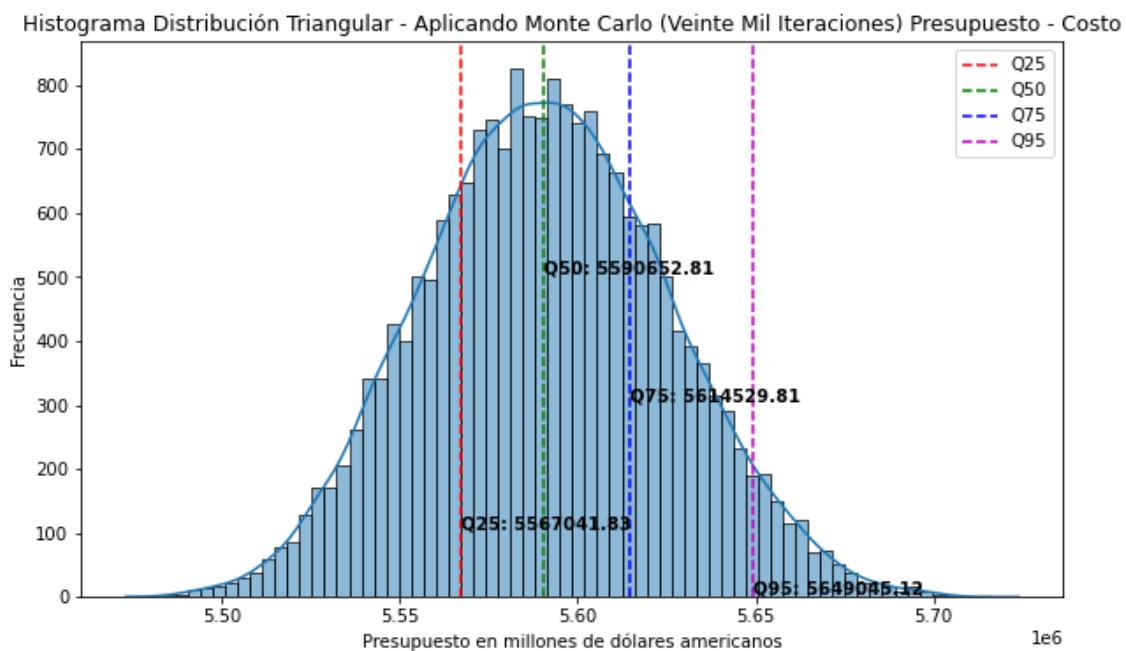


Gráfico 51 Histograma Distribución Triangular – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)

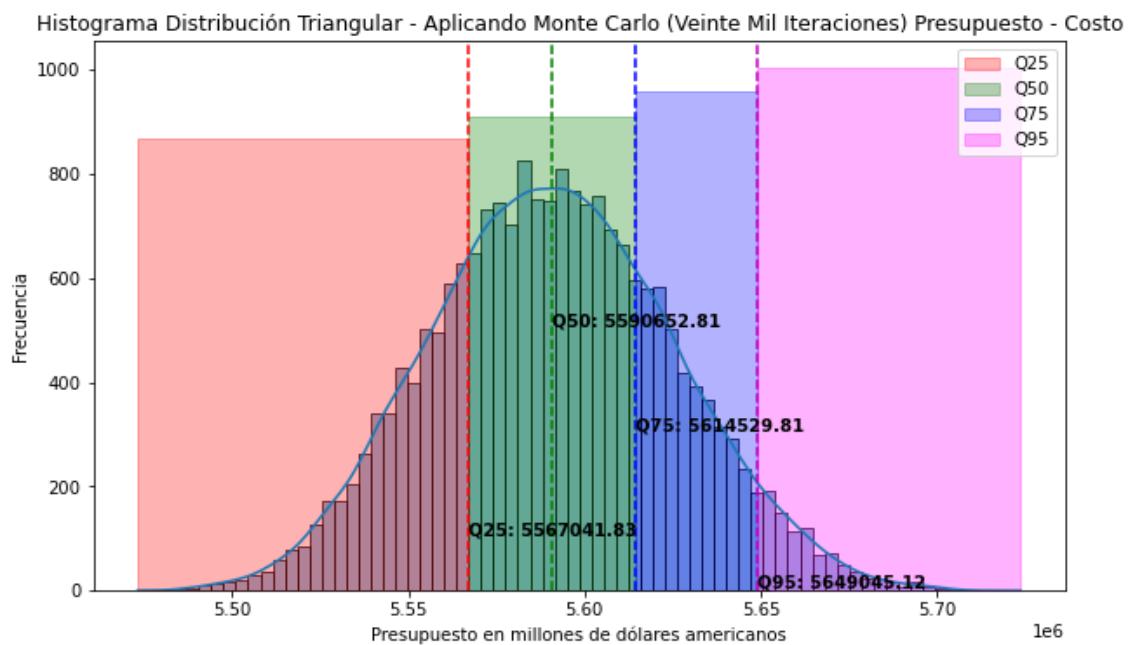
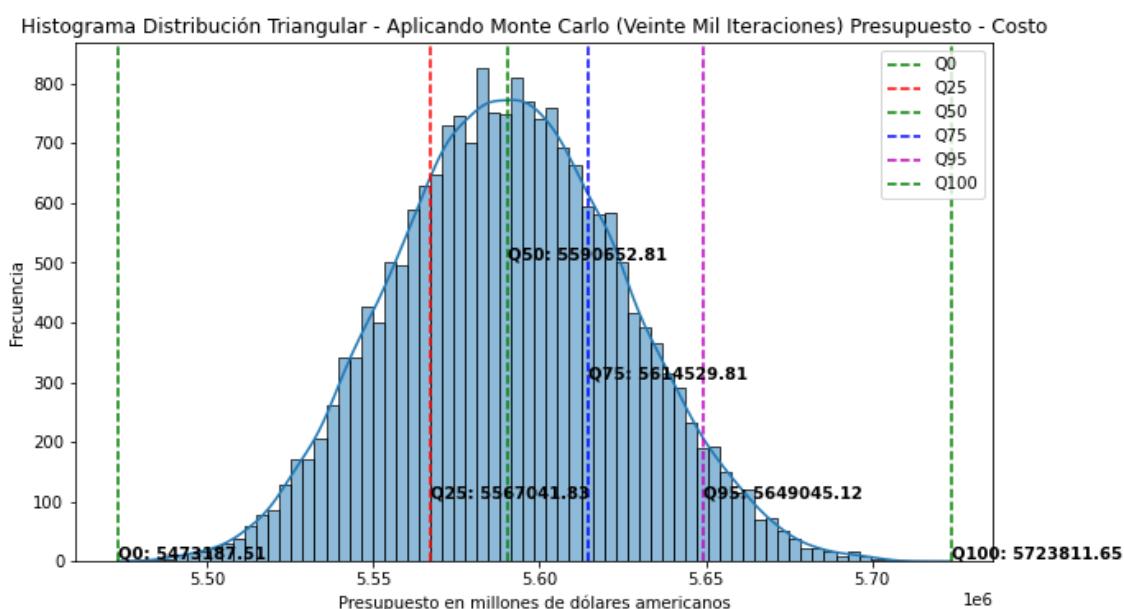


Tabla 18 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000, 10,000 y 20,000 Autor: (Osejo Francisco, 2023)

Distribución Triangular	Simulación Monte Carlo - Iteraciones				Media
	1.000	5.000	10.000	20.000	
Q0	5.480.271,26	5.473.816,12	5.472.634,51	5.473.187,51	5.474.977,35
Q25	5.565.224,64	5.566.873,86	5.567.429,33	5.567.041,83	5.566.642,41
Q50	5.590.160,35	5.589.848,50	5.590.619,61	5.590.652,81	5.590.320,32
Q75	5.614.345,36	5.612.992,37	5.613.817,37	5.614.529,81	5.613.921,23
Q95	5.648.581,97	5.647.432,69	5.647.494,05	5.649.045,12	5.648.138,46
Q100	5.716.498,66	5.715.259,08	5.725.985,25	5.723.811,65	5.720.388,66

La tabla 18 muestra la evolución obtenida en cada conjunto de iteraciones, la precisión va siendo evidente a medida que disminuye la diferencia entre los resultados obtenidos.

Gráfico 52 Histograma Distribución Triangular – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)



Este Gráfico muestra todos los percentiles, lo que, facilita totalmente su análisis.

Tabla 19 Cuadro comparativo de información levantada en cada iteración aplicando monte carlo conjuntamente con la distribución triangular Autor: (Osejo Francisco, 2023)

Distribución Triangular	Simulación Monte Carlo - Iteraciones				Media
	1.000	5.000	10.000	20.000	
Frecuencia	130	330	520	810	447,5
Media	5.590.306,46	5.590.394,50	5.590.869,50	5.591.046,10	5.590.654,14
Desviación estándar	34.774,399	33.801,10	34.156,58	34.547,50	34.319,90
Tiempo ejecución segundos	0,95	1,6	1,58	3,02	1,79
Reserva en dólares	44.571,22	44.659,26	45.134,26	45.310,86	44.918,90
Q0	5.480.271,26	5.473.816,12	5.472.634,51	5.473.187,51	5.474.977,35
Q25	5.565.224,64	5.566.873,86	5.567.429,33	5.567.041,83	5.566.642,41
Q50	5.590.160,35	5.589.848,50	5.590.619,61	5.590.652,81	5.590.320,32
Q75	5.614.345,36	5.612.992,37	5.613.817,37	5.614.529,81	5.613.921,23
Q95	5.648.581,97	5.647.432,69	5.647.494,05	5.649.045,12	5.648.138,46
Q100	5.716.498,66	5.715.259,08	5.725.985,25	5.723.811,65	5.720.388,66
Probabilidad de cumplir el costo objetivo, estimación Más Probable 5.545.735,24 dólares :	9,8%	8,88%	9,83%	9,61%	9,5%
Probabilidad de cumplir el costo objetivo, estimación Optimista 5.324.090,777 dólares :	0%	0%	0%	0%	0%

Probabilidad de cumplir el costo objetivo, estimación Pesimista 5,902,653,451 dólares :	100%	100%	100%	100%	100%
Comprobación Q0,Q25, Q50, Q75, Q95,Q100					
Probabilidad Q0:	0%	0%	0%	0%	0%
Probabilidad Q25:	25%	25%	25%	25%	25%
Probabilidad Q50:	50%	50%	50%	50%	50%
Probabilidad Q75:	75%	75%	75%	75%	75%
Probabilidad Q95:	95%	95%	95%	95%	95%
Probabilidad Q100:	100%	100%	100%	100%	100%

Conclusión:

- Muestra el presupuesto mínimo en millones de dólares **5.474.977,35**, así como la duración máxima en millones de dólares **5.720.388,66**, a más del presupuesto más probable en millones de dólares o Q50 **5.590.320,32**.
- Permite cuantificar el nivel de riesgo existente.
- Facilita determinar qué actividades tienen mayor riesgo
- Suministra información vital para estructurar los respectivos planes de contingencia
- Facilita visualizar las oportunidades – riesgos positivos.
- Ayuda a determinar las reservas de presupuesto - costo
- Obtener y analizar la desviación estándar, por ejemplo:
 - Media de Q50: **5.590.654,14 millones de dólares (+ -) 34.319,90 millones de dólares**
 - Rango posible entre: **5.556.334,1 USD a 5.624.974,04 USD**
- El costo más probable es de **5.590.654,14 USD** que equivale 9,5% de probabilidad de lograr el objetivo, lo que indica que existe alto riesgo, habría que revisar las estimaciones para ver dónde está el error y hacer las correcciones del caso.
- El costo optimista es de **5.324.090,777 USD** que equivale al 0% de probabilidad de lograr el objetivo, lo que indica que existe alto riesgo, por lo tanto, no se lo debería tomar en cuenta.
- El costo pesimista **5,902,653,451 USD** equivale al 100% de probabilidad de lograr el objetivo, pero no es determinante, ya que, la media de Q100 es **5.720.388,66 USD**, por

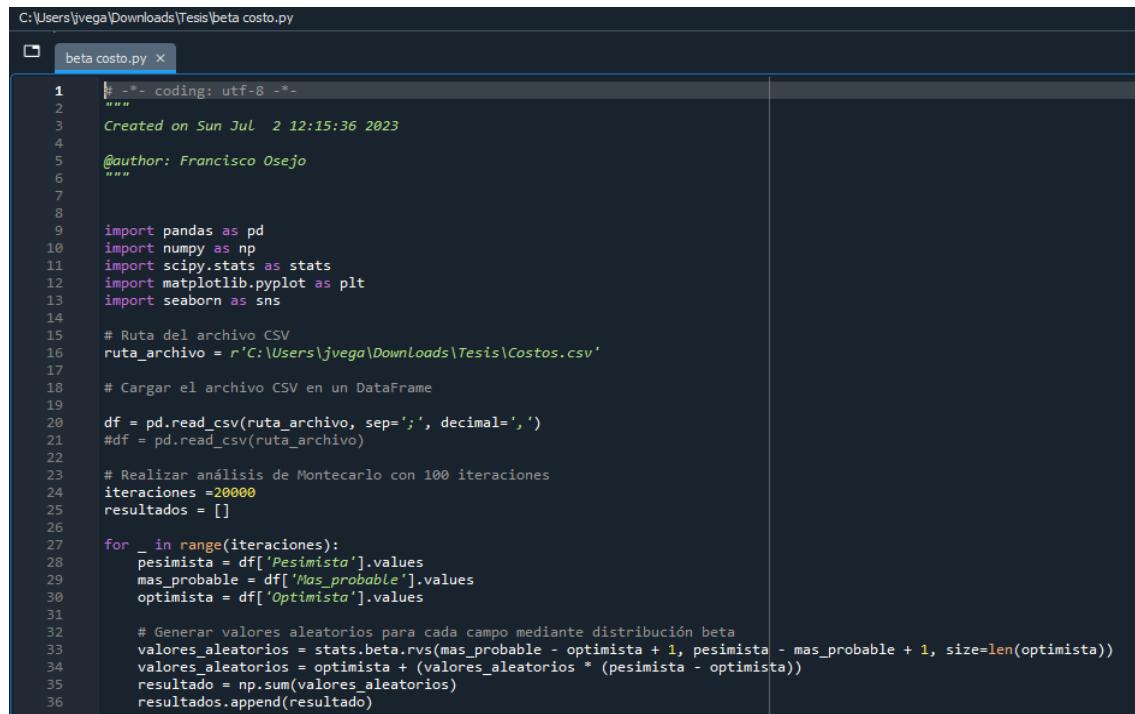
lo tanto, valor pesimista tampoco es buen medidor, lo podríamos considerar un sobreajuste del modelo.

- Por lo tanto, la estimación del presupuesto en millones de sólares hechas en el análisis inicial Vs el resultado de la simulación de Monte Carlo permite determinar que lo estimado inicialmente no es confiable, y se debe volver a reformular en base al resultado obtenido con Monte Carlo.
- Las estimaciones de la duración original Vs la simulación de Montecarlo obtenida determinan que existe un riesgo muy alto de lograr el objetivo, por lo tanto, es recomendable volver revisar los riesgos y reformular las estimaciones de la duración para lograr encontrar el modelo adecuado.
- Es necesario reformular las estimaciones de presupuesto - costo y volver a iterar con Monte Carlo para encontrar el modelo adecuado.

3.5.2.1.2.2 Distribución Probabilística Beta o Pert

3.5.2.1.2.2.1 Código en Python Distribución Probabilística Beta o Pert

Gráfico 53 Simulación Monte Carlo - Código en Python Distribución Beta o Pert
Presupuesto – Costo Autor: (Osejo Francisco, 2023)



The screenshot shows a code editor window with the file 'beta_costo.py' open. The code is written in Python and performs a Monte Carlo simulation for cost estimation. It imports pandas, numpy, scipy.stats, matplotlib.pyplot, and seaborn. It reads a CSV file named 'Costos.csv' located at 'C:\Users\jvega\Downloads\Tesis'. The CSV file has columns 'Pesimista', 'Mas_probable', and 'Optimista'. The script generates 20,000 random values for each column using the beta distribution and calculates the sum of these values for each row, resulting in a new column 'resultado'.

```
C:\Users\jvega\Downloads\Tesis\betaacosto.py
beta costo.py X
1  #-*- coding: utf-8 -*-
2  """
3  Created on Sun Jul  2 12:15:36 2023
4
5  @author: Francisco Osejo
6  """
7
8
9  import pandas as pd
10 import numpy as np
11 import scipy.stats as stats
12 import matplotlib.pyplot as plt
13 import seaborn as sns
14
15 # Ruta del archivo CSV
16 ruta_archivo = r'C:\Users\jvega\Downloads\Tesis\Costos.csv'
17
18 # Cargar el archivo CSV en un DataFrame
19
20 df = pd.read_csv(ruta_archivo, sep=';', decimal=',')
#df = pd.read_csv(ruta_archivo)
22
23 # Realizar análisis de Montecarlo con 100 iteraciones
24 iteraciones = 20000
25 resultados = []
26
27 for _ in range(iteraciones):
28     pesimista = df['Pesimista'].values
29     mas_probable = df['Mas_probable'].values
30     optimista = df['Optimista'].values
31
32     # Generar valores aleatorios para cada campo mediante distribución beta
33     valores_aleatorios = stats.beta.rvs(mas_probable - optimista + 1, pesimista - mas_probable + 1, size=len(optimista))
34     valores_aleatorios = optimista + (valores_aleatorios * (pesimista - optimista))
35     resultado = np.sum(valores_aleatorios)
36     resultados.append(resultado)
37
```

Gráfico 54 Simulación Monte Carlo - Código en Python Distribución Beta o Pert Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\betaacosto.py
beta costo.py x

39     # Calcular los percentiles
40     q0 = np.percentile(resultados, 0)
41     q25 = np.percentile(resultados, 25)
42     q50 = np.percentile(resultados, 50)
43     q75 = np.percentile(resultados, 75)
44     q95 = np.percentile(resultados, 95)
45     q100 = np.percentile(resultados, 100)
46
47     # Imprimir los resultados
48     print("Q0:", q0)
49     print("Q25:", q25)
50     print("Q50 (Mediana):", q50)
51     print("Q75:", q75)
52     print("Q95:", q95)
53     print("Q100:", q100)
54
55     # Crear histograma
56     plt.figure(figsize=(10, 6))
57     sns.histplot(resultados, kde=True)
58
59     # Agregar líneas verticales en los percentiles
60     plt.axvline(q0, color='g', linestyle='--', label='Q0')
61     plt.axvline(q25, color='r', linestyle='--', label='Q25')
62     plt.axvline(q50, color='g', linestyle='--', label='Q50')
63     plt.axvline(q75, color='b', linestyle='--', label='Q75')
64     plt.axvline(q95, color='m', linestyle='--', label='Q95')
65     plt.axvline(q100, color='g', linestyle='--', label='Q100')
66
67     # Agregar anotaciones con los valores de los percentiles en negrita y color negro
68     plt.text(q0, 100, f'Q0: {q0:.2f}', rotation=0, va='bottom', color='black', weight='bold')
69     plt.text(q25, 300, f'Q25: {q25:.2f}', rotation=0, va='bottom', color='black', weight='bold')
70     plt.text(q50, 600, f'Q50: {q50:.2f}', rotation=0, va='bottom', color='black', weight='bold')
71     plt.text(q75, 400, f'Q75: {q75:.2f}', rotation=0, va='bottom', color='black', weight='bold')
72     plt.text(q95, 300, f'Q95: {q95:.2f}', rotation=0, va='bottom', color='black', weight='bold')
73     plt.text(q100, 100, f'Q100: {q100:.2f}', rotation=0, va='bottom', color='black', weight='bold')
74
```

Gráfico 55 Simulación Monte Carlo - Código en Python Distribución Beta o Pert Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\betaacosto.py
beta costo.py x

76     plt.xlabel('Presupuesto en millones de dólares americanos')
77     plt.ylabel('Frecuencia')
78     plt.title('Histograma Distribución Beta o Pert - Aplicando Monte Carlo (Veinte Mil Iteraciones) - Presupuesto - Costo')
79     plt.legend()
80     plt.show()
81
82     # Crear histograma mostrando divisiones de nivel de riesgo
83     plt.figure(figsize=(10, 6))
84     sns.histplot(resultados, kde=True)
85
86     # Rellenar áreas con diferentes colores
87     plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q0, q25, alpha=0.3, color='red', label='Q25')
88     plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q25, q75, alpha=0.3, color='green', label='Q50')
89     plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q75, q95, alpha=0.3, color='blue', label='Q75')
90     plt.fill_betweenx([0, plt.gca().get_ylim()[1]], q95, q100, alpha=0.3, color='magenta', label='Q95')
91
92     # Agregar líneas verticales en los percentiles
93     plt.axvline(q25, color='r', linestyle='--')
94     plt.axvline(q50, color='g', linestyle='--')
95     plt.axvline(q75, color='b', linestyle='--')
96     plt.axvline(q95, color='m', linestyle='--')
97
98     # Agregar anotaciones con los valores de los percentiles en negrita y color negro
99     plt.text(q25, 200, f'Q25: {q25:.2f}', rotation=0, va='bottom', color='black', weight='bold')
100    plt.text(q50, 500, f'Q50: {q50:.2f}', rotation=0, va='bottom', color='black', weight='bold')
101    plt.text(q75, 300, f'Q75: {q75:.2f}', rotation=0, va='bottom', color='black', weight='bold')
102    plt.text(q95, 0, f'Q95: {q95:.2f}', rotation=0, va='bottom', color='black', weight='bold')
103
104    plt.xlabel('Presupuesto en millones de dólares americanos')
105    plt.ylabel('Frecuencia')
106    plt.title('Histograma Distribución Beta o Pert - Aplicando Monte Carlo (Veinte Mil Iteraciones) - Presupuesto - Costo')
107    plt.legend()
108    plt.show()
```

Gráfico 56 Simulación Monte Carlo - Código en Python Distribución Beta o Pert Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
109     # Calcular la media y la desviación estándar de los resultados
110     media = np.mean(resultados)
111     desviacion_estandar = np.std(resultados)
112
113
114     # Calcular la probabilidad de cumplir el presupuesto objetivo
115     # Presupuesto objetivo
116     presupuesto_objetivo = 5545735.24
117     probabilidad_cumplir = sum(resultado <= presupuesto_objetivo for resultado in resultados) / iteraciones
118
119     print("Probabilidad de cumplir el presupuesto objetivo:", probabilidad_cumplir)
120
```

Se realizó la simulación de Monte Carlo para 1.000, 5.000, 10.000 y 20.000 iteraciones aplicando stats.beta.rvs () de scipy de manera independiente.

```
# Generar valores aleatorios para cada campo mediante distribución beta

valores_aleatorios = stats.beta.rvs(mas_probable - optimista + 1, pesimista -
mas_probable + 1, size=len(optimista))

valores_aleatorios = optimista + (valores_aleatorios * (pesimista - optimista))

resultado = np.sum(valores_aleatorios)

resultados.append(resultado) (Osejo Francisco, 2023).
```

Para efectos de este trabajo de titulación sólo mostraremos el código aplicado a las 20,000 iteraciones.

Para las 1.000, 5.000, 10.000 iteraciones hay que cambiar en el código el valor de la variable “iteraciones =”, adicionalmente cambiar posiciones y textos en el código correspondiente a la generación de los histogramas respectivos en cada ejecución del programa. (Osejo Francisco, 2023).

3.5.2.1.2.2.2 Histogramas obtenidos en Python Distribución Probabilística Beta o Pert

Para efectos de este trabajo de titulación solo mostraremos Histogramas obtenidos en las simulaciones de 10,000 y 20,000 iteraciones

Gráfico 57 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Diez Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)

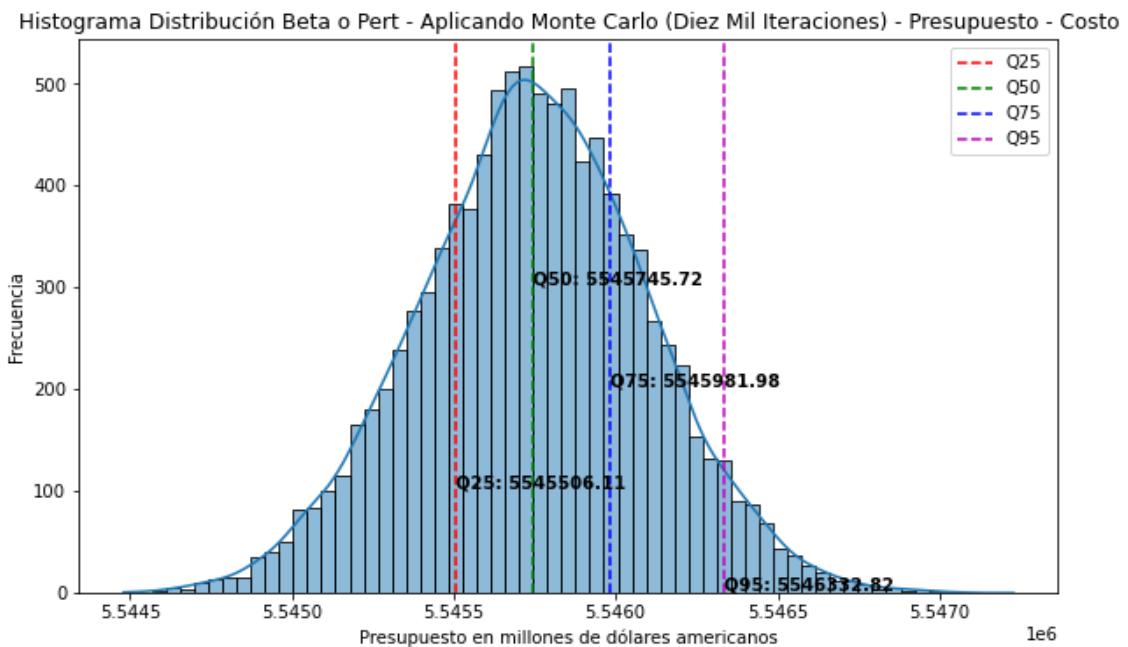
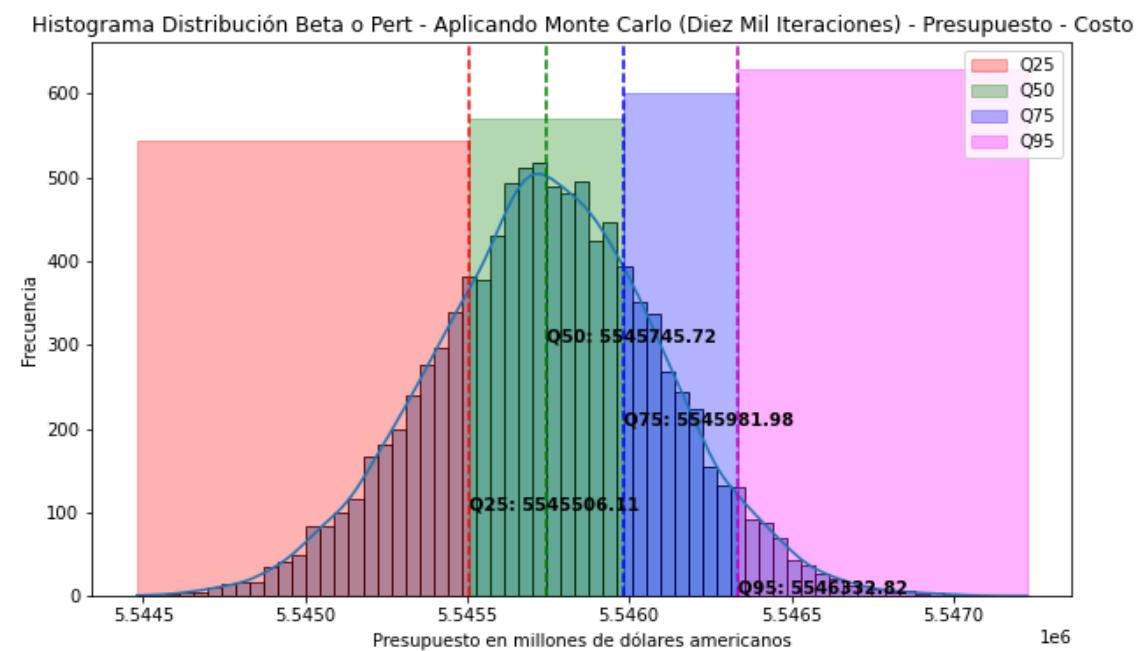


Gráfico 58 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Diez Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)



Como podemos observar en los Gráficos 57 y 58 de este documento, el código del algoritmo realiza 10,000 diez mil iteraciones en donde aplicamos stats.beta.rvs () de la librería scipy

```
valores_aleatorios = stats.beta.rvs(mas_probable - optimista + 1, pesimista - mas_probable + 1, size=len(optimista)) (Osejo Francisco, 2023)
```

Realiza los cálculos en base los valores generados randomicamente posibles dentro de los rangos de los valores de las variables optimista y pesimista (Osejo Francisco, 2023).

Obtenemos los percentiles Q0, Q25, Q50, Q75, Q95, Q100:

Tabla 20 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000 y 10,000 Autor: (Osejo Francisco, 2023)

Distribución Beta-Pert	Simulación Monte Carlo - Iteraciones		
	1.000	5.000	10.000
Q0	5.544.353,934	5.544.491,931	5.544.479,629
Q25	5.545.495,525	5.545.503,17	5.545.506,11
Q50	5.545.727,164	5.545.749,48	5.545.745,72
Q75	5.545.987,268	5.545.988,32	5.545.981,98
Q95	5.546.333,879	5.546.319,91	5.546.332,82
Q100	5.547.040,730	5.547.158,898	5.547.228,655

Q0: Muestra el presupuesto mínimo en millones de dólares que pueden iniciar las actividades

Q50: Muestra el presupuesto medio en millones de dólares en los cuales las actividades pueden ejecutarse

Q100: Muestra el presupuesto máximo en millones de dólares que pueden terminar las actividades

La interpretación de los Gráficos 57 y 58 es la siguiente:

- Permite identificar visualmente los límites de los rangos de tiempos en días
- Facilita el análisis de estimaciones de tiempo en días

- Muestra con mayor precisión las probabilidades de lograr el objetivo
 - Q0: Cualquier presupuesto menor o igual al Q0 tendrá 0% de probabilidades de lograr el objetivo.
 - Q25: Cualquier presupuesto menor o igual al Q25 tendrá de 0% a 25% de probabilidades de lograr el objetivo.
 - Q50: Cualquier presupuesto menor o igual al Q50 tendrá de 26% a 50% de probabilidades de lograr el objetivo.
 - Q75: Cualquier presupuesto menor o igual al Q75 tendrá de 51% a 75% de probabilidades de lograr el objetivo.
 - Q100: Cualquier presupuesto menor o igual al Q100 tendrá de 76% a 100% de probabilidades de lograr el objetivo.
- Obtener los percentiles nos permite determinar el nivel de riesgo existente en el cumplimiento de los objetivos.
 - Q25: Indica que existe un 25% de probabilidad de éxito, lo que implica un alto riesgo de lograr el objetivo.
 - Q50: Indica que existe un 50% de probabilidad de éxito, lo que implica un riesgo medio de lograr el objetivo.
 - Q75: Indica que existe un 75% de probabilidad de éxito, lo que implica un riesgo bajo de lograr el objetivo.
 - Q100: Indica que existe un 100% de probabilidad de éxito, lo que implica que no existe riesgo de lograr el objetivo.
- Conseguir las ponderaciones de riesgo (Alto, medio, bajo) conjuntamente con las probabilidades de éxito permite determinar la matriz de probabilidad e impacto con esquema de puntuación que se muestra en el Gráfico 5
- La Frecuencia muestra claramente el nivel de iteraciones que coincidieron en ese rango de análisis. Por lo que, es muy importante revisar los resultados obtenidos en cada conjunto de iteraciones.
- La información descrita en la Tabla 20 muestra que a mayor número de iteraciones (1000, 5000, 10000) se logra afinar la precisión en los resultados obtenidos.

Gráfico 59 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)

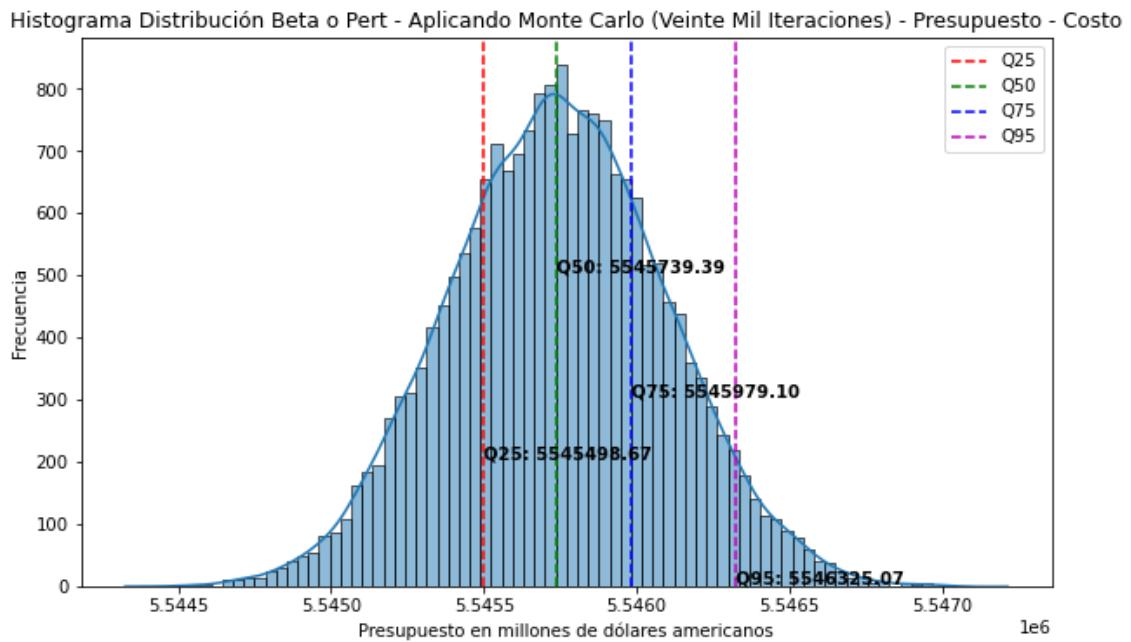


Gráfico 60 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)

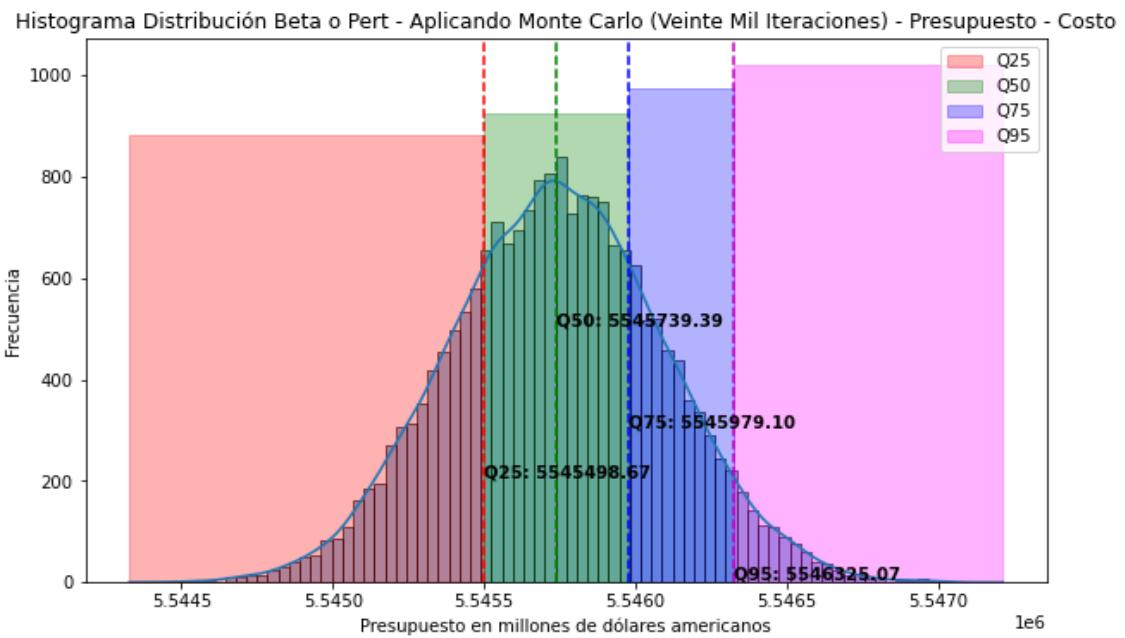
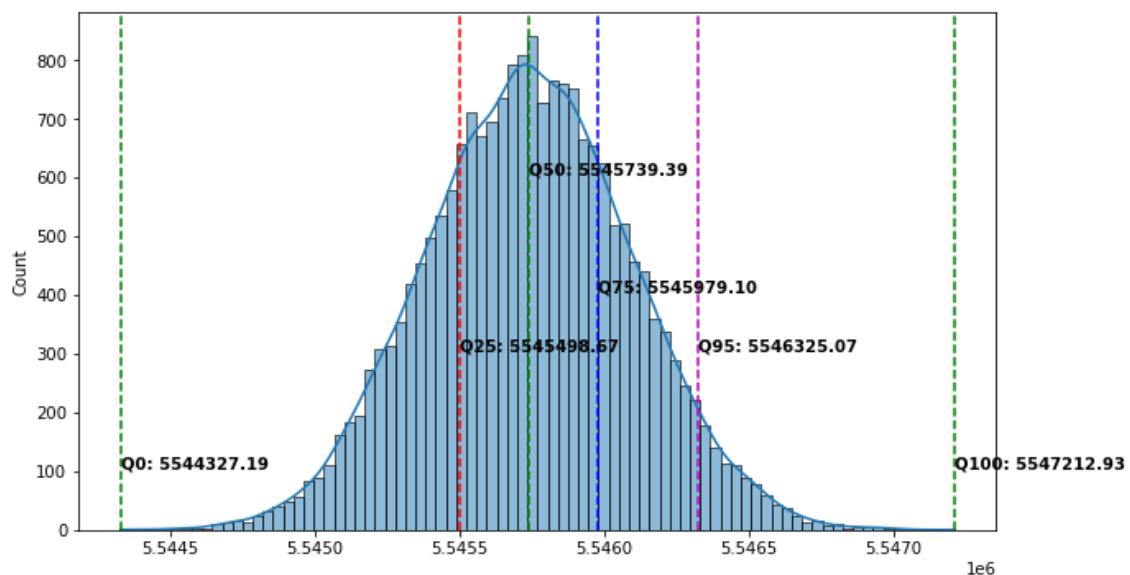


Tabla 21 Tiempos en días obtenidos en percentiles calculados en las iteraciones de 1,000, 5,000, 10,000 y 20,000 Autor: (Osejo Francisco, 2023)

Distribución Beta-Pert	Simulación Monte Carlo - Iteraciones				Media
	1.000	5.000	10.000	20.000	
Q0	5.544.353,934	5.544.491,931	5.544.479,629	5.544.327,19	5.544.413,170
Q25	5.545.495,525	5.545.503,17	5.545.506,11	5.545.498,67	5545500,868
Q50	5.545.727,164	5.545.749,48	5.545.745,72	5.545.739,39	5545740,438
Q75	5.545.987,268	5.545.988,32	5.545.981,98	5.545.979,10	5545984,168
Q95	5.546.333,879	5.546.319,91	5.546.332,82	5.546.325,07	5546327,919
Q100	5.547.040,730	5.547.158,898	5.547.228,655	5.547.212,93	5.547.160,30

La tabla 21 muestra la evolución obtenida en cada conjunto de iteraciones, la precisión va siendo evidente a medida que disminuye la diferencia entre los resultados obtenidos.

Gráfico 61 Histograma Distribución Beta o Pert – Aplicando Monte Carlo (Veinte Mil iteraciones) – Presupuesto – Costo Autor: (Osejo Francisco, 2023)



Este Gráfico muestra todos los percentiles, lo que, facilita totalmente su análisis.

Tabla 22 Cuadro comparativo de información levantada en cada iteración aplicando monte carlo conjuntamente con la distribución beta o pert Autor: (Osejo Francisco, 2023)

Distribución Beta-Pert	Simulación Monte Carlo - Iteraciones				Media
	1.000	5.000	10.000	20.000	
Frecuencia	115	340	550	850	463,75
Media	5.545.736,81	5.545.744,034	5.545.744,31	5.545.739,22	5.545.741,09
Desviación estándar	363,86	355,53	357,50	355,82	358,179
Tiempo ejecución segundos	1,12	2,02	4,05	4,60	2,948
Reserva en dólares	1,57	8,79	9,07	3,98	5,853
Q0	5.544.353,934	5.544.491,93	5.544.479,63	5.544.327,19	5.544.413,170
Q25	5.545.495,525	5.545.503,17	5.545.506,11	5.545.498,67	5545500,868
Q50	5.545.727,164	5.545.749,48	5.545.745,72	5.545.739,39	5545740,438
Q75	5.545.987,268	5.545.988,32	5.545.981,98	5.545.979,10	5545984,168
Q95	5.546.333,879	5.546.319,91	5.546.332,82	5.546.325,07	5546327,919
Q100	5.547.040,730	5.547.158,90	5.547.228,66	5.547.212,93	5547160,304
Probabilidad de cumplir el costo objetivo, estimación Más Probable 5.545.735,24 dólares :	51%	48%	48,80%	49,45%	49,31%
Probabilidad de cumplir el costo objetivo, estimación Optimista 5.324.090,777 dólares :	0%	0%	0%	0%	0%
Probabilidad de cumplir el costo objetivo, estimación Pesimista 5,902,653,451 dólares :	100%	100%	100%	100%	100%
Comprobación Q0,Q25, Q50, Q75, Q95,Q100					
Probabilidad Q0:	0,00%	0,00%	0,00%	0,00%	0,00%
Probabilidad Q25:	25%	25%	25%	25%	25,00%
Probabilidad Q50:	50%	50%	50%	50%	50,00%
Probabilidad Q75:	75%	75%	75%	75%	75,00%
Probabilidad Q95:	95%	95%	95%	95%	95,00%
Probabilidad Q100:	100%	100%	100%	100%	100,00%

Conclusión:

- Muestra presupuesto mínimo en **USD 5.544.413,170**, así como el presupuesto máximo en **USD 5.547.160,304** a más del presupuesto más probable en USD o Q50 **5.545.741,09**.
- Permite cuantificar el nivel de riesgo existente.
- Facilita determinar qué actividades tienen mayor riesgo
- Suministra información vital para estructurar los respectivos planes de contingencia
- Facilita visualizar las oportunidades – riesgos positivos.
- Ayuda a determinar las reservas de tiempo
- Obtener y analizar la desviación estándar, por ejemplo:
 - Media de Q50: **5.545.741,09 USD (+ -) 358,179 USD**
 - Rango posible entre: **5.545.382.92 USD a 5.546.099,27 USD**
- El valor más probable es de **5.545.741,09 USD** que equivale **49,39%** de probabilidad de lograr el objetivo, lo que indica que existe riesgo medio, habría que revisar las estimaciones para ver dónde está el error y hacer las correcciones del caso.
- El valor optimista es de **5.324.090,777 USD** que equivale al 0% de probabilidad de lograr el objetivo, lo que indica que existe alto riesgo, por lo tanto, no se lo debería tomar en cuenta.
- El valor pesimista es de **5,902,653,451 USD** que equivale al 100% de probabilidad de lograr el objetivo, pero no es determinante, ya que, la media de Q100 es **5547160,304 USD**, por lo tanto, valor pesimista tampoco es buen medidor, lo podríamos considerar un sobreajuste del modelo.
- Por lo tanto, la estimación del presupuesto en millones de dólares hecha en el análisis inicial Vs el resultado de la simulación de Monte Carlo permite determinar que lo estimado inicialmente no es confiable, y se debe volver a reformular en base al resultado obtenido con Monte Carlo.
- Las estimaciones del presupuesto original Vs la simulación de Montecarlo obtenida determinan que existe un riesgo muy alto de lograr el objetivo, por lo tanto, es recomendable volver revisar los riesgos y reformular las estimaciones de la duración para lograr encontrar el modelo adecuado.

- Es necesario reformular las estimaciones del presupuesto y volver a iterar con Monte Carlo para encontrar el modelo adecuado.

3.5.2.2 Regresión Lineal

La Regresión lineal se explica con mayor detalle en la sección 2.10 de este trabajo de titulación.

Dataset Cronograma – Tiempo. - luego de realizar los laboratorios pertinentes se determinó que no es aplicable realizar regresión lineal, ya que existe alta relación entre las variables y esto genera multicolinealidad.

Este modelo tiene por objetivo realizar lo siguiente:

- a) Aplicar Regresión Lineal Simple
- b) Aplicar Regresión Lineal Múltiple
- c) Usar el dataset datasets de presupuesto – costo de forma separada para aplicar tanto la regresión lineal simple, así como la regresión lineal múltiple respectivamente.
- d) Aplicar la regresión lineal simple y múltiple haciendo uso de la librería SKlearn.
- e) Realizar Cross Validation con el 80% de la data para entrenamiento y 20% de la data para prueba.
- f) Seleccionar las variables independientes necesarias para pronosticar el costo – presupuesto.
- g) Realizar el análisis de coeficientes
- h) Documentar el código en Python
- i) Graficar
- j) Registrar los resultados.
- k) Conclusiones

3.5.2.2.1 Aplicar Regresión Lineal Simple a Dataset de Presupuesto – Costo

Gráfico 62 Regresión Lineal Simple Dataset Presupuesto – Costo Autor: (Osejo Francisco, 2023)

Index	Descripcion	Tiempo	Optimista	as probal	Pesimista	ucion Tri	tribucion I
0	FIDS/BIDS/RMS/AODB/CUTE - Lumpsum (Ultra) ->	292.88	982070	984171	986272	984171	984171
1	Servers y SO->	27	78549	92410.6	104424	91794.5	92102.6
2	Workstation (HP) ->	45	78213.8	85015	96917.1	86715.3	85865.1
3	DDC (ACER) ->	72	73320	78000	92040	81120	79560
4	Monitores 42->	71	96360	109500	122640	109500	109500
5	Brackets for FIDS Monitors + Imstallation ->	65	44500	50000	58000	50833.3	50416.7
6	LAN and VoIP - Materiales (Cisco / DOS) ->	194	974763	975713	983734	978070	976892
7	LAN and VoIP - Servicios (Cisco / DOS) ->	444.63	190144	208950	231934	210342	209646
8	Master Clock (Moba Time) ->	230	9310.95	10120.6	11638.7	10356.7	10238.7
9	FIDS Labor JCI ->	188.88	323921	344597	406625	358381	351489
10	FIDS FMT JCI ->	64	286668	329503	382223	332798	331151
11	Security Systems: Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	287	975718	983394	995238	984783	984088
12	Subcontract Installation ->	89	66750	75000	85500	75750	75375
13	Security Labor JCI ->	104	190000	200000	226000	205333	202667
14	Security FMT JCI ->	50	106814	113632	123859	114768	114200
15	Public Address (Bosh) ->	90	353400	380000	399000	377467	378733
16	Ingenieria PA (Bosh) ->	30	31850	35000	39200	35350	35175
17	Notifier ->	120	246298	256560	289913	264257	260408
18	Wattstopper ->	180	37600	40000	45600	41066.7	40533.3
19	PA Labor JCI ->	110	105411	114578	127181	115723	115150
20	PA FMT JCI ->	90	72428.7	79592	94714.5	82245.1	80918.5

3.5.2.2.1.1 Código en Python Regresión Lineal Simple

Gráfico 63 Regresión Lineal Simple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)

The screenshot shows a code editor window with the following details:

- File path: C:\Users\jvega\Downloads\Tesis\Regresion Lineal Simple Costo.py
- Tab title: Regresion Lineal Simple Costo.py
- Code content:

```
1  # -*- coding: utf-8 -*-
2  """
3      Created on Sat Jul 01 21:54:50 2023
4
5      @author: Francisco Osejo
6  """
7
8  import pandas as pd
9  from sklearn.model_selection import train_test_split
10 from sklearn.linear_model import LinearRegression
11 import matplotlib.pyplot as plt
12
13 # Ruta del archivo CSV
14 ruta_archivo = r'C:\Users\jvega\Downloads\Tesis\Costos.csv'
15
16 # Cargar el archivo CSV en un DataFrame
17
18 df = pd.read_csv(ruta_archivo, sep=';', decimal=',')
19
20 # Dividir los datos en variables de entrada (X) y variable objetivo (y)
21 X = df[['Tiempo']]
22 y = df['Mas_probable']
23
24 # Grafiquemos los datos correspondientes
25
26 plt.scatter(X,y)
27 plt.xlabel('Tiempo en días')
28 plt.ylabel('Costo medio')
29 plt.title('Regresión lineal')
30 plt.legend()
31 plt.show()
```

Gráfico 64 Regresión Lineal Simple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\Regresion Lineal Simple Costo.py
Regresion Lineal Simple Costo.py x

32
33 # Dividir los datos en conjuntos de entrenamiento y prueba (80% - 20%)
34 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
35
36 # Crear un modelo de regresión lineal y entrenarlo
37 model = LinearRegression()
38 model.fit(X_train, y_train)
39
40 # Realizar predicciones en el conjunto de prueba
41 y_pred = model.predict(X_test)
42
43 # Crear un gráfico de dispersión con la línea de regresión
44 plt.scatter(X_test['Tiempo'], y_test, color='blue', label='Datos reales')
45 plt.plot(X_test['Tiempo'], y_pred, color='red', linewidth=2, label='Línea de regresión')
46 plt.xlabel('Tiempo')
47 plt.ylabel('Valor medio')
48 plt.title('Regresión Lineal')
49 plt.legend()
50 plt.show()
51
52 ...
53
54 Calcular los valores de la pendiente y la intersección para ver
55 como queda la ecuación del modelo, la Regresión Lineal Simple
56 viene representada como y= ax+b, por lo tanto
57 ...
58 print()
59 print('DATOS DEL MODELO REGRESIÓN LINEAL SIMPLE')
60 print()
61 print('Valor de la pendiente o coeficiente "a": ')
62 print(model.coef_)
63 print('Valor de la intersección o coeficiente "b": ')
64 print(model.intercept_)
65 print()
66 print('La ecuación del modelo es igual a: ')
67 print ('y=', model.coef_, 'X', model.intercept_)

68
```

Gráfico 65 Regresión Lineal Simple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
53 ...
54 Calcular los valores de la pendiente y la intersección para ver
55 como queda la ecuación del modelo, la Regresión Lineal Simple
56 viene representada como y= ax+b, por lo tanto
57 ...
58 print()
59 print('DATOS DEL MODELO REGRESIÓN LINEAL SIMPLE')
60 print()
61 print('Valor de la pendiente o coeficiente "a": ')
62 print(model.coef_)
63 print('Valor de la intersección o coeficiente "b": ')
64 print(model.intercept_)
65 print()
66 print('La ecuación del modelo es igual a: ')
67 print ('y=', model.coef_, 'X', model.intercept_)

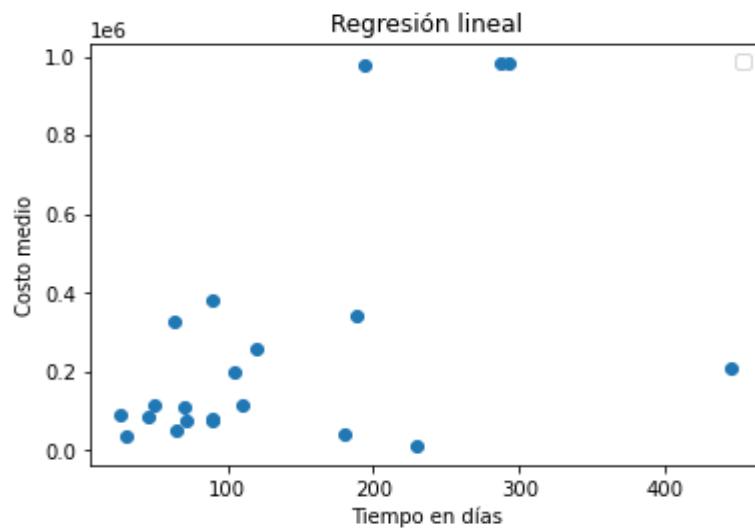
68
69 # Calcular el coeficiente de determinación (R^2)
70 r2_score = model.score(X_test, y_test)
71
72 # Imprimir el coeficiente de determinación
73 print("Coeficiente de determinación (R^2):", r2_score)
74
75 ...
76 Como ya lo habíamos visto comparando los resultados de "y_predict" y viendo la gráfica podíamos
77 deducir que la precisión no iba a ser muy buena
78 ...
```

En este modelo de regresión lineal simple se aplica al dataset de presupuesto – costo, se va a predecir el presupuesto – costo en base a duración – tiempo de cada actividad inmersa en el dataset (Osejo Francisco, 2023).

Variable dependiente = Mas_probable (costo más probable de la estimación original)

Variable independiente = Tiempo

Gráfico 66 Regresión Lineal Simple – Gráfico de dispersión Presupuesto – Costo Autor: (Osejo Francisco, 2023)



El Gráfico 66 muestra los datos utilizando una gráfica de dispersión para ver exactamente la distribución de los datos. A simple vista ya se puede deducir que el modelo de Regresión Lineal Simple que vamos a construir tendrá un error muy alto (Osejo Francisco, 2023).

Gráfico 67 Regresión Lineal Simple – Datos y_test – Costo Autor: (Osejo Francisco, 2023)

Index	Mas probable
0	984171
17	256560
15	380000
1	92410.6
8	10120.6

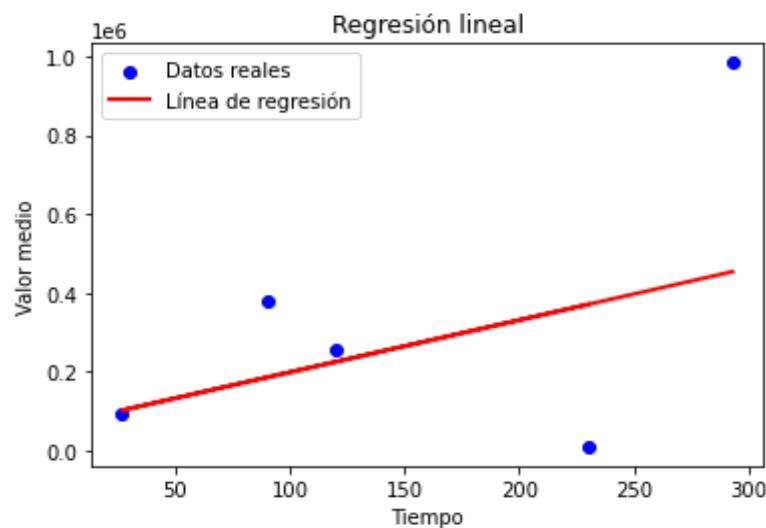
Gráfico 68 Regresión Lineal Simple – Datos y_predit – Costo Autor: (Osejo Francisco, 2023)

0	453954
0	225306
1	185628
3	102306
4	370790

Al comparar con los valores obtenidos en nuestra predicción con los valores reales, para ver que tal es el comportamiento de nuestro modelo. Como podemos observar en el Gráfico 67 y_test tenemos los datos reales mientras que en Gráfico 68 y_pred están los datos obtenidos de nuestra predicción implementando el modelo. Si observamos con

detenimiento estos datos nos podemos dar cuenta que en ciertos momentos los datos reales son muy parecidos con los datos que se han calculado, mientras que otros casos los datos son totalmente distintos, Con esta pequeña comparación podemos determinar que nuestro modelo no es del todo eficiente (Osejo Francisco, 2023).

Gráfico 69 Regresión Lineal Simple – Gráfico de dispersión con la línea de regresión –
Costo Autor: (Osejo Francisco, 2023)



Como ya se lo había visto comparando los resultados de "y_predict" y viendo la gráfica 69 se puede deducir que la precisión no iba a ser muy buena. (Osejo Francisco, 2023)

Gráficamente se puede observar la línea roja sería nuestro modelo y los puntos azules son los datos con los que se entrenó. Si se revisa el objetivo de este algoritmo es que el modelo o línea abarque el mayor número de muestras y de esta forma disminuir el error que vendría siendo la distancia desde la línea hasta el punto, entonces todos los datos fuera de la línea serían errores, es por esta razón que cuando se realiza la comparación de los datos, anteriormente, se encuentra que varios tenían una diferencia considerable, bueno la razón de esto es que el modelo no abarcó todos los puntos. De igual forma cuando se grafican los datos primeramente pudimos observar que no se iba a obtener buenos resultados con este algoritmo, y la razón es muy simple, los datos se encuentran muy dispersos, por lo que, es difícil que la línea cubriera la mayoría de los datos, por lo que, se iba a obtener una mala precisión, esto no quiere decir este algoritmo sea malo

sino más bien que este algoritmo no es bueno para el conjunto de datos – dataset (Osejo Francisco, 2023).

El Gráfico 65 muestra el código en Python que genera la siguiente información:

- Cálculo del coeficiente de determinación R ^2
- Valor de la pendiente o coeficiente “a”
- Valor del interceptor o coeficiente “b”
- La ecuación del modelo

Gráfico 70 Regresión Lineal Simple – Datos del modelo - Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
Console 1/A ×  
DATOS DEL MODELO REGRESIÓN LINEAL SIMPLE  
Valor de la pendiente o coeficiente "a":  
[1322.58500455]  
Valor de la intersección o coeficiente "b":  
66595.71013500963  
La ecuación del modelo es igual a:  
y= [1322.58500455] X 66595.71013500963  
Coeficiente de determinación (R^2): 0.24171249134560446  
In [12]:
```

Ecuación del modelo es igual a:

$$Y = [1322.585] X 66595,710$$

Coeficiente de determinación R ^2 = 0.24177

El valor obtenido en el coeficiente de determinación R^2 indica la precisión del algoritmo, como ya se pudo comparar los resultados de “y_predit” y viendo el Gráfico 66 es posible deducir que la precisión del modelo no iba a ser muy buena, el valor obtenido de R^2 ratifica que el modelo como tal no es muy preciso, esto no significa que el algoritmo sea malo, sino que no es el mejor para este conjunto de datos – dataset (Osejo Francisco, 2023).

3.5.2.2.2 Aplicar Regresión Lineal Múltiple al Dataset de Presupuesto – Costo

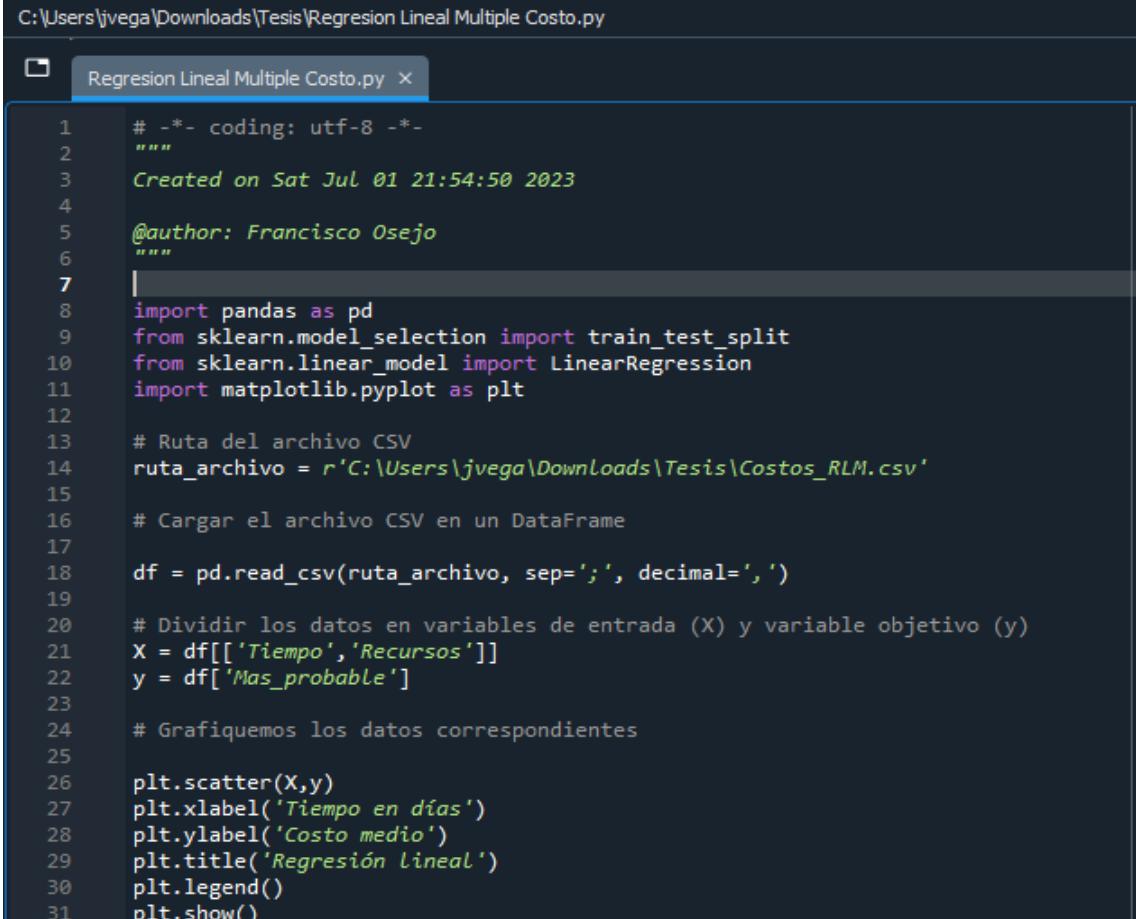
Para poder aplicar Regresión Lineal Múltiple hemos incorporado la variable recursos al Dataset original de presupuesto – costo.

Gráfico 71 Regresión Lineal Múltiple Dataset Presupuesto – Costo Autor: (Osejo Francisco, 2023)

Index	Descripción	Tiempo	Recursos	Otimista	Mas probable	Pesimista	Distribución Triangular	Distribución beta
0	FIDS/BIDS/RMS/AOB/CUTE - Lumpsum (Ultra) ->	292.88	28	982070	984171	986272	984171	984171
1	Servers y 50->	27	5	78549	92418.6	104424	91794.5	92102.6
2	Workstation (MP) ->	45	10	78213.8	85015	96917.1	86715.3	85865.1
3	DOC (ACER) ->	72	10	73320	78000	92048	81120	79568
4	Monitores 42->	71	10	96368	109500	122640	109500	109500
5	Brackets for FIDS Monitors + Installation ->	65	10	44500	50000	58000	50833.3	50416.7
6	LAN and VoIP - Materiales (Cisco / DOS) ->	194	5	974763	975713	983734	978070	976892
7	LAN and VoIP - Servicios (Cisco / DOS) ->	444.63	15	198144	208950	231934	210342	209646
8	Master Clock (Moba Time) ->	230	2	9310.95	10120.6	11638.7	10356.8	10238.7
9	FIDS Labor JCI ->	188.88	10	323921	344597	406625	358381	351489
10	FIDS FMT JCI ->	64	5	286668	329503	382223	332798	331151
11	Security Systems: Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	287	20	975718	983394	995238	984783	984088
12	Subcontract Installation ->	89	15	66750	75000	85500	75750	75375
13	Security Labor JCI ->	104	20	190000	200000	226000	205333	202667
14	Security FMT JCI ->	50	5	106814	113632	123859	114768	114200
15	Public Address (Bosh) ->	90	10	353400	380000	399000	377467	378733
16	Ingeniería PA (Bosh) ->	30	10	31850	35000	39200	35350	35175
17	Notifier ->	120	5	246298	256560	289913	264257	260408
18	Wattstopper ->	180	5	37600	40000	45600	41066.7	40533.3
19	PA Labor JCI ->	110	10	105411	114578	127181	115723	115150
20	PA FMT JCI ->	90	10	72428.7	79592	94714.5	82245.1	80918.5

3.5.2.2.2.1 Código en Python Regresión Lineal Múltiple

Gráfico 72 Regresión Lineal Múltiple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)



The screenshot shows a code editor window with the following details:

- Path: C:\Users\jvega\Downloads\Tesis\Regresion Lineal Multiple Costo.py
- Title bar: Regresion Lineal Multiple Costo.py
- Code content:

```
1  # -*- coding: utf-8 -*-
2  """
3      Created on Sat Jul 01 21:54:50 2023
4
5      @author: Francisco Osejo
6      """
7
8  import pandas as pd
9  from sklearn.model_selection import train_test_split
10 from sklearn.linear_model import LinearRegression
11 import matplotlib.pyplot as plt
12
13 # Ruta del archivo CSV
14 ruta_archivo = r'C:\Users\jvega\Downloads\Tesis\Costos_RLM.csv'
15
16 # Cargar el archivo CSV en un DataFrame
17
18 df = pd.read_csv(ruta_archivo, sep=';', decimal=',')
19
20 # Dividir los datos en variables de entrada (X) y variable objetivo (y)
21 X = df[['Tiempo', 'Recursos']]
22 y = df['Mas_probable']
23
24 # Grafiquemos los datos correspondientes
25
26 plt.scatter(X,y)
27 plt.xlabel('Tiempo en días')
28 plt.ylabel('Costo medio')
29 plt.title('Regresión Lineal ')
30 plt.legend()
31 plt.show()
```

Gráfico 73 Regresión Lineal Múltiple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
C:\Users\jvega\Downloads\Tesis\Regresion Lineal Multiple Costo.py
  □ Regresion Lineal Multiple Costo.py ×

32
33     # Dividir los datos en conjuntos de entrenamiento y prueba (80% - 20%)
34     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
35
36     # Crear un modelo de regresión lineal y entrenarlo
37     model = LinearRegression()
38     model.fit(X_train, y_train)
39
40     # Realizar predicciones en el conjunto de prueba
41     y_pred = model.predict(X_test)
42
43     # Crear un gráfico de dispersión con la línea de regresión
44     plt.scatter(X_test['Tiempo'], y_test, color='blue', label='Datos reales')
45     plt.plot(X_test['Tiempo'], y_pred, color='red', linewidth=2, label='Línea de regresión')
46     plt.xlabel('Tiempo')
47     plt.ylabel('Valor medio')
48     plt.title('Regresión Lineal')
49     plt.legend()
50     plt.show()
51
52
```

Gráfico 74 Regresión Lineal Múltiple - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
53 """
54 Calcular los valores de la pendiente y la intersección para ver
55 como queda la ecuación del modelo, la Regresión Lineal Simple
56 viene representada como  $y = ax + b$ , por lo tanto
57 """
58 print()
59 print('DATOS DEL MODELO REGRESIÓN LINEAL SIMPLE')
60 print()
61 print('Valor de la pendiente o coeficiente "a": ')
62 print(model.coef_)
63 print('Valor de la intersección o coeficiente "b": ')
64 print(model.intercept_)
65 print()
66 print('La ecuación del modelo es igual a: ')
67 print('y =', model.coef_, 'X', model.intercept_)
68
69 # Calcular el coeficiente de determinación ( $R^2$ )
70 r2_score = model.score(X_test, y_test)
71
72 # Imprimir el coeficiente de determinación
73 print("Coeficiente de determinación ( $R^2$ ):", r2_score)
74
75 print('Precisión del modelo: ')
76 print(model.score(X_train, y_train))
77
78 ...
80 Como ya lo habíamos visto comparando los resultados de "y_predict" y viendo la gráfica podíamos
81 deducir que la precisión no iba a ser muy buena
82 """
```

En este modelo de regresión lineal múltiple se aplica al dataset de presupuesto - costo, se va a predecir el presupuesto en base al tiempo y el número de recursos de cada actividad inmersa en el dataset (Osejo Francisco, 2023).

Variable dependiente = Mas_probable (costo más probable de la estimación original)

Variable independiente 1 = Tiempo

Variable independiente 2 = Recursos

Gráfico 75 Regresión Lineal Múltiple – Datos y_test – Costo Autor: (Osejo Francisco, 2023)

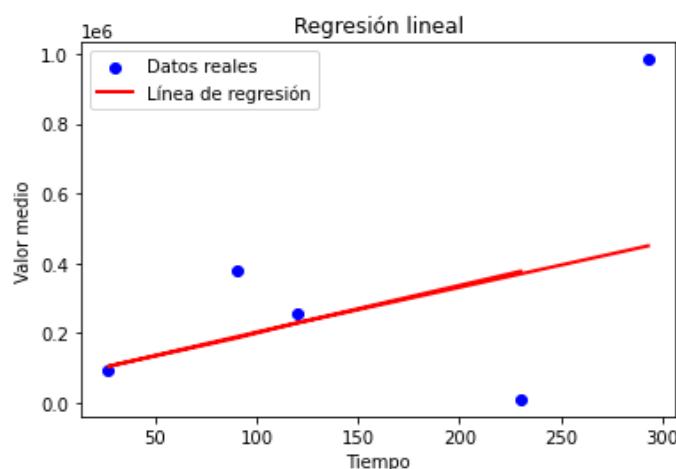
Index	Mas_probable
0	984171
17	256560
15	380000
1	92410.6
8	10120.6

Gráfico 76 Regresión Lineal Múltiple – Datos y_pred – Costo Autor: (Osejo Francisco, 2023).

y_pred - NumPy object array	
0	449879
1	228592
2	185616
3	104696
4	376942

Al comparar con los valores obtenidos en nuestra predicción con los valores reales, para ver que tal es el comportamiento de nuestro modelo. Como podemos observar en el Gráfico 75 y_test tenemos los datos reales mientras que en Gráfico 76 y_pred están los datos obtenidos de nuestra predicción implementando el modelo. Si observamos con detenimiento estos datos nos podemos dar cuenta que en ciertos momentos los datos reales son muy parecidos con los datos que se han calculado, mientras que otros casos los datos son totalmente distintos. Con esta pequeña comparación podemos determinar que nuestro modelo no es del todo eficiente (Osejo Francisco, 2023).

Gráfico 77 Regresión Lineal Múltiple – Gráfico de dispersión con la línea de regresión – Costo Autor: (Osejo Francisco, 2023)



Como ya se lo había visto comparando los resultados de "y_predict" se puede deducir que la precisión no iba a ser muy buena. (Osejo Francisco, 2023)

Gráficamente se puede observar la línea roja sería nuestro modelo y los puntos azules son los datos con los que se entrenó. Si se revisa el objetivo de este algoritmo es que el modelo o línea abarque el mayor número de muestras y de esta forma disminuir el error que vendría siendo la distancia desde la línea hasta el punto, entonces todos los datos fuera de la línea serían errores, es por esta razón que cuando se realiza la comparación de los datos, anteriormente, se encuentra que varios tenían una diferencia considerable, bueno la razón de esto es que el modelo no abarcó todos los puntos. De igual forma cuando se grafican los datos primeramente pudimos observar que no se iba a obtener buenos resultados con este algoritmo, y la razón es muy simple, los datos se encuentran muy dispersos, por lo que, es difícil que la línea cubriera la mayoría de los datos, por lo que, se iba a obtener una mala precisión, esto no quiere decir este algoritmo sea malo sino más bien que este algoritmo no es bueno para el conjunto de datos – dataset (Osejo Francisco, 2023).

El Gráfico 74 muestra el código en Python que genera la siguiente información:

- Cálculo del coeficiente de determinación R ^2
- Valor de la pendiente o coeficiente “a”
- Valor del interceptor o coeficiente “b”
- La ecuación del modelo

Gráfico 78 Regresión Lineal Múltiple – Datos del modelo - Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
DATOS DEL MODELO REGRESIÓN LINEAL SIMPLE

Valor de la pendiente o coeficiente "a":
[1332.22005504 -601.85904782]
Valor de la intersección o coeficiente "b":
71735.19007457767

La ecuación del modelo es igual a:
y= [1332.22005504 -601.85904782] X 71735.19007457767
Coeficiente de determinación (R^2): 0.2270929581016684
Precisión del modelo:
0.22368834047017017

In [23]:
```

Ecuación del modelo es igual a:

$$Y = [1332.2200] X - 601,859$$

Coeficiente de determinación R² = 0.2236

El valor obtenido en el coeficiente de determinación R² indica la precisión del algoritmo, como ya se pudo comparar los resultados de “y_predit” es posible deducir que la precisión del modelo no iba a ser muy buena, el valor obtenido de R² ratifica que el modelo como tal no es muy preciso a pesar de haber aumentado una nueva variable independiente al modelo, esto no significa que el algoritmo sea malo, sino que no es el mejor para este conjunto de datos – dataset.

3.5.2.3 Arboles de Decisión

Arboles de Decisión se explica con mayor detalle en la sección 2.9 de este trabajo de titulación.

Dataset Cronograma – Tiempo. - luego de realizar los laboratorios pertinentes se determinó que no es aplicable realizar arboles de decisión, ya que existe alta relación entre las variables y esto genera multicolinealidad.

Este modelo tiene por objetivo realizar lo siguiente:

- a) Aplicar Árbol de Decisión
- b) Usar el dataset de presupuesto – costo
- c) Aplicar Árbol de Decisión haciendo uso de la librería SKlearn.
- d) Realizar Cross Validation con el 80% de la data para entrenamiento y 20% de la data para prueba.
- e) Seleccionar las variables independientes necesarias para pronosticar el costo – presupuesto.
- f) Realizar el análisis de coeficientes
- g) Documentar el código en Python
- h) Graficar
- i) Registrar los resultados.
- j) Conclusiones

Gráfico 79 Arboles de Decisión Dataset Presupuesto – Costo Autor: (Osejo Francisco, 2023)

df - DataFrame

Index	Descripción	Tiempo	Optimista	Mas probable	Pesimista	Ucion	Tribucion	L
0	FIDS/BIDS/RMS/AODB/CUTE - Lumpsum (Ultra) ->	292.88	982070	984171	986272	984171	984171	
1	Servers y SO->	27	78549	92410.6	104424	91794.5	92102.6	
2	Workstation (HP) ->	45	78213.8	85015	96917.1	86715.3	85865.1	
3	DDC (ACER) ->	72	73320	78000	92040	81120	79560	
4	Monitores 42->	71	96360	109500	122640	109500	109500	
5	Brackets for FIDS Monitors + Imstallation ->	65	44500	50000	58000	50833.3	50416.7	
6	LAN and VoIP - Materiales (Cisco / DOS) ->	194	974763	975713	983734	978070	976892	
7	LAN and VoIP - Servicios (Cisco / DOS) ->	444.63	190144	208950	231934	210342	209646	
8	Master Clock (Moba Time) ->	230	9310.95	10120.6	11638.7	10356.7	10238.7	
9	FIDS Labor JCI ->	188.88	323921	344597	406625	358381	351489	
10	FIDS FMT JCI ->	64	286668	329503	382223	332798	331151	
11	Security Systems: Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	287	975718	983394	995238	984783	984088	
12	Subcontract Installation ->	89	66750	75000	85500	75750	75375	
13	Security Labor JCI ->	104	190000	200000	226000	205333	202667	
14	Security FMT JCI ->	50	106814	113632	123859	114768	114200	
15	Public Address (Bosh) ->	90	353400	380000	399000	377467	378733	
16	Ingenieria PA (Bosh) ->	30	31850	35000	39200	35350	35175	
17	Notifier ->	120	246298	256560	289913	264257	260408	
18	Wattstopper ->	180	37600	40000	45600	41066.7	40533.3	
19	PA Labor JCI ->	110	105411	114578	127181	115723	115150	
20	PA FMT JCI ->	90	72428.7	79592	94714.5	82245.1	80918.5	

3.5.2.3.1 Código en Python Regresión Lineal Múltiple

Gráfico 80 Arboles de Decisión - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
# -*- coding: utf-8 -*-
"""
Created on Sat Jul 01 21:54:50 2023
@author: Francisco Osejo
"""

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor, plot_tree
import matplotlib.pyplot as plt

# Ruta del archivo CSV
ruta_archivo = r'C:\Users\jvega\Downloads\Tesis\Costos.csv'

# Cargar el archivo CSV en un DataFrame
df = pd.read_csv(ruta_archivo, sep=';', decimal=',')

# Dividir los datos en variables de entrada (X) y variable objetivo (y)
X = df[['Tiempo']]
y = df['Mas_probable']

# Graficamos los datos correspondientes
plt.scatter(X,y)
plt.show()

# Dividir los datos en conjuntos de entrenamiento y prueba (80% - 20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Defino el algoritmo a utilizar
adr = DecisionTreeRegressor(max_depth = 5)
```

Gráfico 81 Árboles de Decisión - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
57     # Entreno el modelo
58
59     adr.fit(X_train, y_train)
60
61     # Realizo la predicción
62
63     y_pred = adr.predict(X_test)
64
65
66
67     print('DATOS DEL MODELO ÁRBOLES DE DECISIÓN REGRESIÓN')
68     print()
69     print('Precision del modelo: ')
70     print(adr.score(X_train,y_train))
71
72
73     # Graficamos los datos de prueba junto con la predicción
74
75     # Generar x_grid para la línea de predicción
76     x_grid = np.arange(min(X_test.values), max(X_test.values), 0.1)
77     x_grid = x_grid.reshape(len(x_grid), 1)
78
79     # Graficar los puntos de prueba y la línea de predicción
80     plt.scatter(X_test, y_test, label='Datos de prueba')
81     plt.plot(x_grid, adr.predict(x_grid), color='red', linewidth=3, label='Predicción')
82     plt.xlabel('Tiempo')
83     plt.ylabel('Mas_probable')
84     plt.title('Árbol de Decisión - Predicción')
85     plt.legend()
86     plt.grid(True)
87     plt.show()
```

Gráfico 82 Árboles de Decisión - Código en Python Presupuesto – Costo Autor: (Osejo Francisco, 2023)

```
70     # Graficar el árbol de decisión
71     plt.figure(figsize=(12, 8))
72     plot_tree(adr, filled=True, feature_names=['Tiempo'], rounded=True, fontsize=10)
73     plt.title('Árbol de decisión')
74     plt.show()
75
76
77     # Graficar el árbol de decisión
78     plt.figure(figsize=(14, 10))
79     plot_tree(adr, filled=True, feature_names=['Tiempo'], rounded=True, fontsize=12, label='all', proportion=True, precision=2, impurity=False)
80     plt.title('Árbol de decisión', fontsize=16)
81     plt.show()
82
83
84     # Crear una figura y ejes para el gráfico del árbol
85     plt.figure(figsize=(14, 10))
86     ax = plt.subplot(1, 1, 1)
87
88     # Graficar el árbol de decisión con detalles mejorados
89     plot_tree(adr, filled=True, feature_names=['Tiempo'], rounded=True, fontsize=12, precision=2, impurity=False, ax=ax)
90
91     # Personalizar la apariencia del árbol
92     ax.set_title('Árbol de Decisión', fontsize=18)
93     ax.set_xlabel('Tiempo', fontsize=14)
94     ax.set_ylabel('Mas_probable', fontsize=14)
95     ax.grid(True)
96     plt.tight_layout()
97
98     # Mostrar el gráfico
99     plt.show()
```

En este modelo de árboles de decisión se aplica al dataset de presupuesto - costo, se va a predecir el presupuesto en base al tiempo de cada actividad inmersa en el dataset (Osejo Francisco, 2023).

Variable dependiente = *Mas_probable* (costo más probable de la estimación original)

Variable independiente = *Tiempo*

Gráfico 83 Arboles de Decisión – Datos y_test – Costo Autor: (Osejo Francisco, 2023)

Index	as probal
0	984171
17	256560
15	380000
1	92410.6
8	10120.6

Gráfico 84 Arboles de Decisión – Datos y_pred – Costo Autor: (Osejo Francisco, 2023)

0	
0	983394
1	157289
2	78418.4
3	35000
4	975713

Como se puede observar en el Gráfico 83 se muestran los datos reales y_test mientras que en el Gráfico 84 y_pred están los datos obtenidos de la predicción implementando el modelo. Si observamos con detenimiento estos datos nos podemos dar cuenta que en ciertos momentos los datos reales son muy parecidos con los datos que se han calculado, mientras que otros casos los datos son totalmente distintos (Osejo Francisco, 2023).

Si graficamos los datos tenemos el siguiente Gráfico 85 es un conjunto de datos que se encuentra distribuido uniforme en su mayoría, claro está existen datos dispersos, pero en esta ocasión son mínimos (Osejo Francisco, 2023).

Gráfico 85 Arboles de Decisión – Gráfico de dispersión – Costo Autor : (Osejo Francisco, 2023)

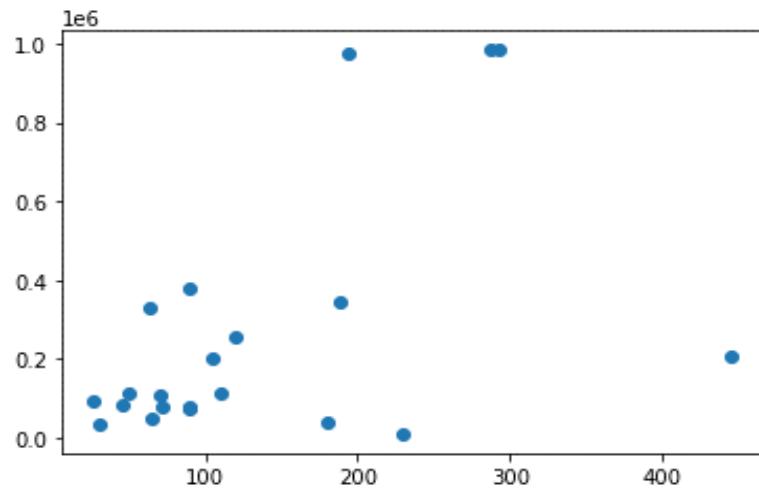


Gráfico 86 Arboles de Decisión – Predicción – Costo Autor: (Osejo Francisco, 2023)

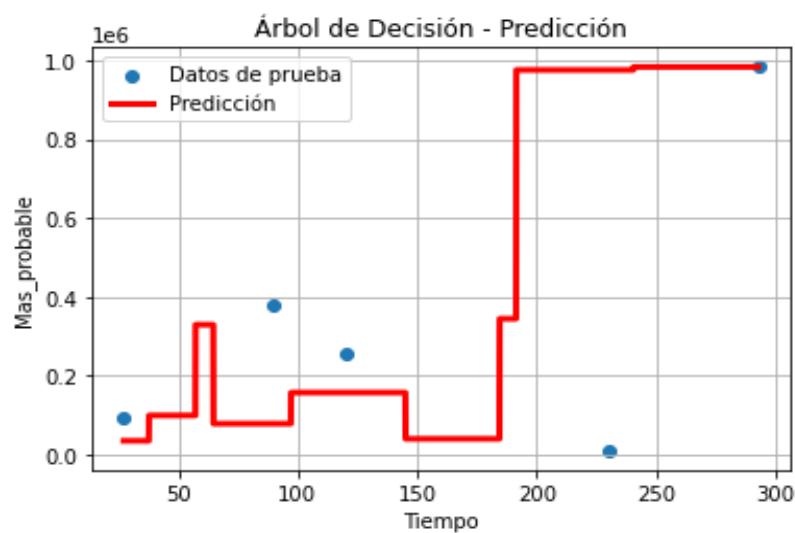


Gráfico 86 permite visualizar los datos de entrenamiento junto con el modelo y su respectiva predicción, como podemos observar el modelo no es una línea recta, como se lo vio en los algoritmos de regresión lineal simple y múltiple, acá se puede observar que el modelo cuenta con subidas y bajadas (Osejo Francisco, 2023).

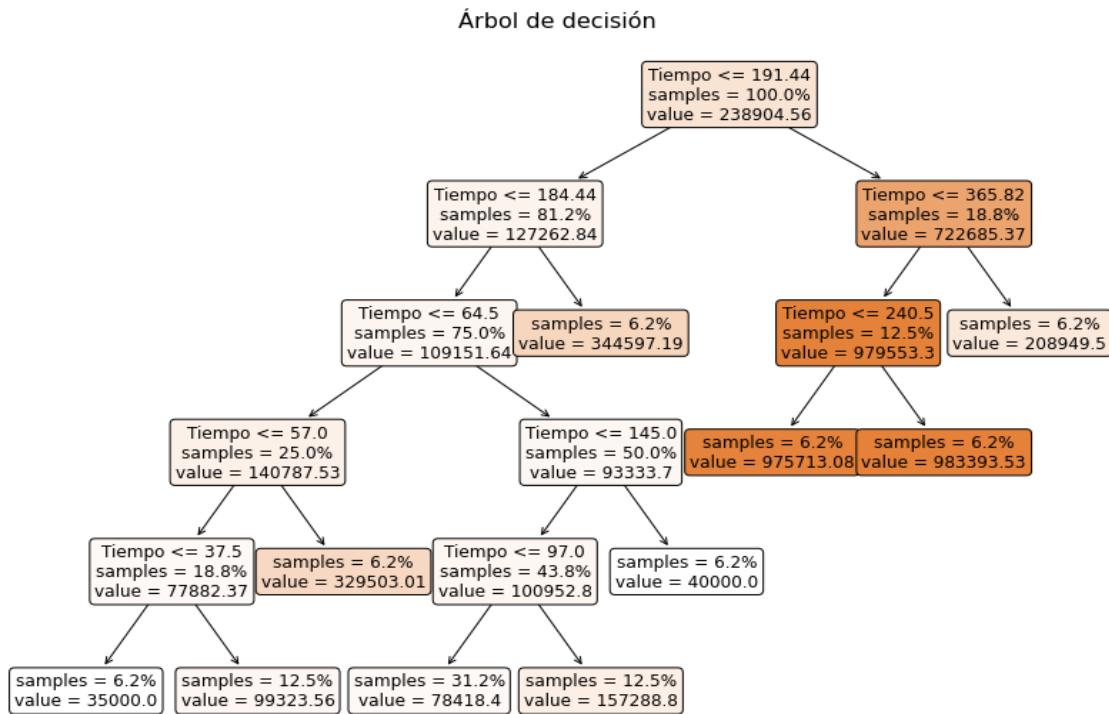
Gráfico 87 Arboles de Decisión – Precisión del Modelo – Costo Autor: (Osejo Francisco, 2023)

```
In [3]: print('DATOS DEL MODELO ARBOLES DE DECISIÓN REGRESIÓN')
...: print()
...: print('Precision del modelo: ')
...: print(adr.score(X_train,y_train))
DATOS DEL MODELO ARBOLES DE DECISIÓN REGRESIÓN

Precision del modelo:
0.9957813344638511
```

El Gráfico 87 muestra la precisión del modelo utilizando la instrucción score el resultado obtenido es de 0,99578 valor muy cercano a 1 lo que implica una excelente precisión del modelo. En este caso Arboles de Decisión Regresión se aplica perfectamente al dataset en cuestión (Osejo Francisco, 2023).

Gráfico 88 Árboles de Decisión con porcentaje de frecuencia - Presupuesto – Costo
Autor: (Osejo Francisco, 2023)



El código en Python de este Gráfico se encuentra detallado en el Gráfico 82 de este documento.

Gráfico 88 muestra un árbol de decisión que está parametrizado a 5 niveles como máximo. Describe didácticamente el comportamiento del modelo, se busca predecir la variable dependiente que en este caso es el costo – presupuesto, a través del comportamiento de la variable independiente la cual es el tiempo – duración de ejecución de la actividad, como tal. Es evidente la correlación existente entre tiempo y costo y sus posibles variaciones como; a mayor tiempo mayor costo, o a bajo presupuesto menor tiempo (Osejo Francisco, 2023).

La construcción del árbol de decisión parte de la media de la variable independiente – tiempo, para este caso la media corresponde a 191,44 días.

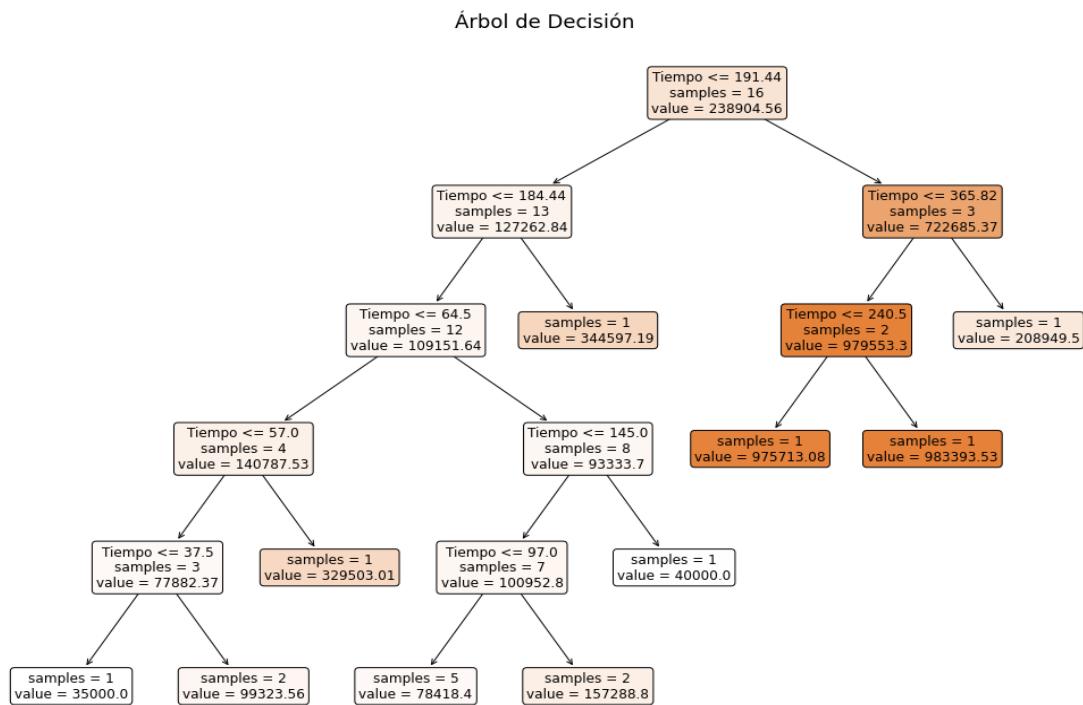
En el primer ramal la pregunta es tiempo $\leq 191,44$ días, 100% de la muestra y la predicción del costo igual a \$238.904,56, la pregunta tiene dos opciones Verdadero o Falso. En este caso la opción de Falso nos lleva por más ramales con sus valores en días porcentaje de frecuencias y predicciones de costos, información importante que nos

permite realizar los análisis pertinentes de riesgo positivo y/o negativos, pero no es lo que estamos buscando.

Al momento de analizar la pregunta del primer ramal y siendo la respuesta Verdadera, vemos que la pregunta del segundo ramal es: tiempo \leq 365,82 días, si la respuesta es Verdadera existe un posible impacto del 6,2% de riesgo positivos 0 negativos, por definir, pero al momento en que la respuesta es Negativa se genera un tercer ramal que cuestiona si el Tiempo \leq 240,5 días con un 12% y una predicción de costo de \$ 979.553,3, como podemos ver ya sea la respuesta Verdadera o Falsa nos lleva a un frecuencia del 6,2% con una predicción de costo de \$975.713,08 y \$ 983.393,53 respectivamente. Lo anteriormente analizado determina que a este ramal debemos darle la mayor atención del caso, ya que, el impacto es muy alto en cuestión de presupuesto – costo, y por ende existe alto riesgo de no lograr el objetivo del proyecto (Osejo Francisco, 2023).

Siguiente paso sería comprobar estos resultados con los obtenidos en el análisis de simulación de Monte Carlo, lo que permitiría determinar las posibles desviaciones entre los valores optimistas, más probables y pesimistas, dando como resultado una mejor percepción de las acciones a tomar (Osejo Francisco, 2023).

Gráfico 89 Árboles de Decisión con frecuencia - Presupuesto – Costo Autor: (Osejo Francisco, 2023)



El código en Python de este Gráfico se encuentra detallado en el Gráfico 82 de este documento.

Gráfico 89 nos muestra a diferencia del Gráfico 89 la frecuencia en cantidad y no en porcentaje.

El análisis y conclusiones son las mismas detalladas en la explicación del Gráfico 89.

3.6 Evaluación

3.6.1 Evaluación de Resultados.

En esta etapa se formaliza su evaluación de si los resultados cumplen o no con los criterios de éxito comercial. Este paso requiere una comprensión clara de los objetivos comerciales establecidos, así que asegúrese de incluir a los tomadores de decisiones clave en la evaluación del proyecto (CRISP-DM, 2021).

3.6.1.1 Evaluación de Dataset de Cronograma – Tiempo.

En el capítulo 4 de este documento se determina que el mejor modelo para este Dataset es Monte Carlo junto con la Distribución Beta Pert, por lo que, el siguiente análisis se basa en los datos obtenidos (Osejo Francisco, 2023).

Gráfico 90 Histograma Distribución Beta - Cronograma – Tiempo Autor: (Osejo Francisco, 2023)

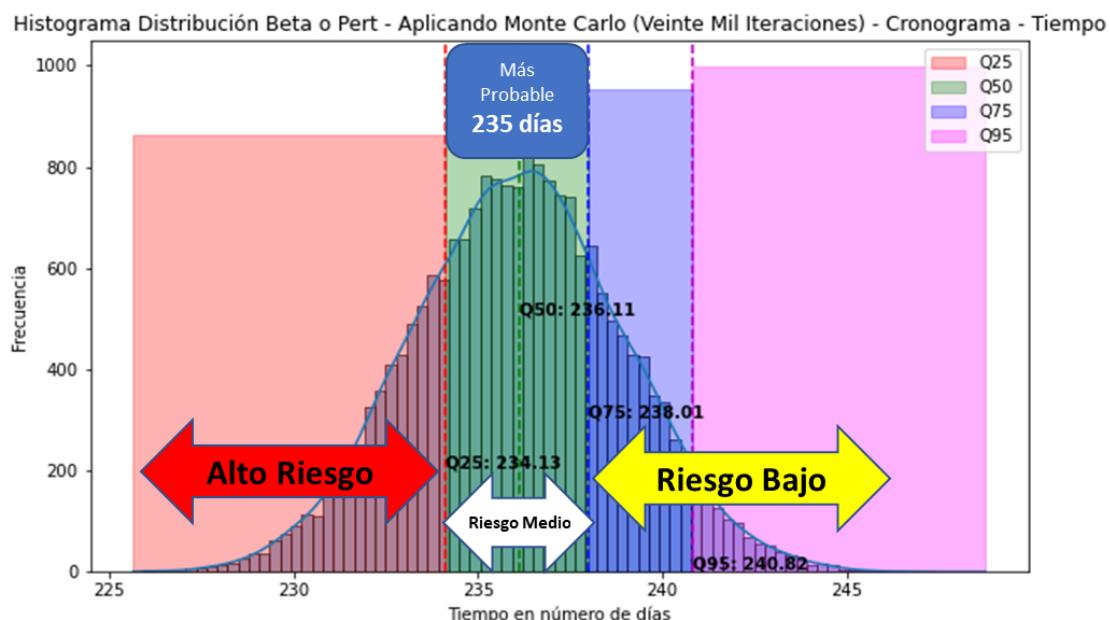


Tabla 23 Análisis estimación más probable original Autor: (Osejo Francisco, 2023)

#	Actividades	Optimista	Mas_probable	Pesimista
	AODB / FIDS / BIDS / GM / CUTE Installation	207,28	235,00	268,50
1	Installation and Configuration	61,00	64,00	67,00
2	Install and configurate servers MCR1	3,00	4,00	6,00
3	Install and configurate servers MCR2	2,00	4,00	6,00
4	Update system configurations	8,00	10,00	12,00
5	Airlines systems/infos. Ready	1,50	2,00	3,50
6	Configure airline interfaces	15,00	17,00	18,00
7	Install Display's Mountings/Brackets	18,00	20,00	22,00
8	Install LCDs & confirm operational	14,00	15,00	17,00
9	Configure DDC	18,00	20,00	22,00
10	Install DDC	9,00	10,00	12,00

11	Integration of DDC with Monitor and servers	8,00	10,00	13,00
12	Monitors and DDC are done	12,00	15,00	17,00
13	SiAT for FIDS	7,00	8,00	9,00
14	set up of CUSE	9,00	10,00	14,00
15	Install CUSE peripherals	8,00	10,00	12,00
16	Integration CUSE at locations	13,00	15,00	16,00
17	Site Acceptance Test (SiAT) for CUSE	0,78	1,00	2,00

La tabla 23 muestra la estimación más probable original con su valor de 235 días, mientras que el Gráfico 91 muestra los percentiles obtenidos en la simulación de Monte Carlo con 20 mil iteraciones, en este Gráfico se puede apreciar los niveles de riesgo existente, los mismos que son: Riesgo alto, riesgo medio y riesgo bajo.

La duración más probable de la tabla 23 es 235 días y la probabilidad de lograr el objetivo es de 34,72% Tabla 27, y se encuentra en la zona de riesgo medio según Gráfico 91.

El análisis pertinente para este caso es revisar los riesgos identificados de las actividades 1,7, y 9 de la tabla 23 que son los que muestran mayor duración.

Se observa que la reserva de tiempo es de 1,13 días según la tabla 27, lo que determina que el modelo es muy óptimo.

La desviación estándar del modelo es (+-) 2,872 días según la tabla 27 indica que la dispersión en torno a la media es baja, lo que ratifica que existe un riesgo medio.

Se recomienda en base a los resultados obtenidos en el modelo, volver a estimar las duraciones actividades, queda a libertad del equipo ir hacia la izquierda o derecha de la media, ya dependerá si son adversos o arriesgados al riesgo (Osejo Francisco, 2023).

3.6.1.2 Evaluación de Dataset de Presupuesto – Costo.

En el capítulo 4 de este documento se determina que hay dos modelos que se aplican perfectamente al Dataset de Presupuesto – Costo, los cuales son:

- Simulación Monte Carlo aplicando Distribución Beta – Pert
- Arboles de Decisiones

3.6.1.2.1 Simulación Monte Carlo aplicando Distribución Beta – Pert

Gráfico 91 Histograma Distribución Beta - Presupuesto – Costo Autor: (Osejo Francisco, 2023)

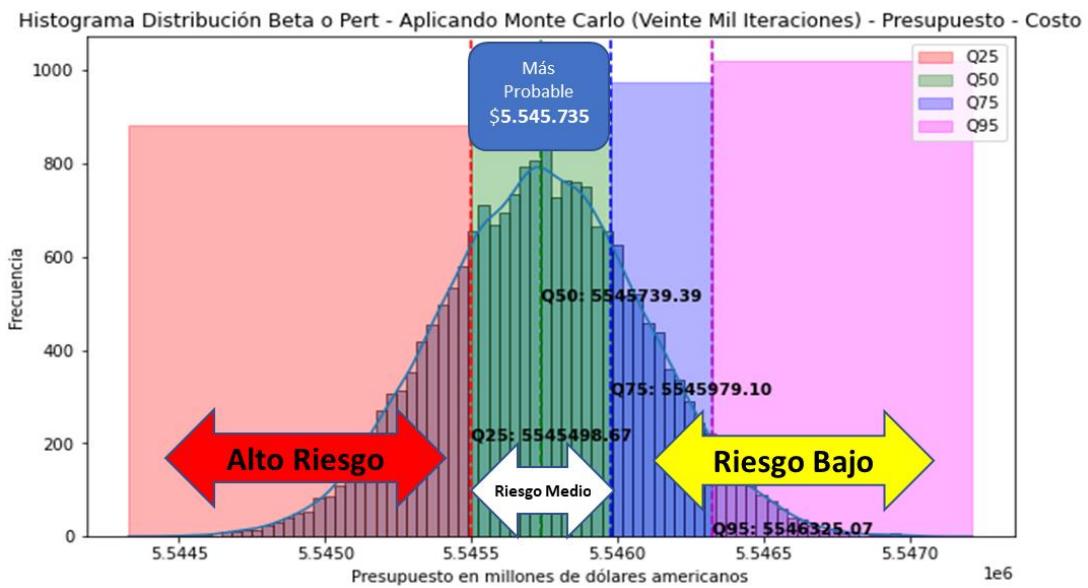


Tabla 24 Análisis estimación más probable original - presupuesto – costo Autor: (Osejo Francisco, 2023)

Actividad	Tiempo	Recursos	Optimista	Mas_probable	Pesimista
FIDS/BIDS/RMS/AODB/CUTE - Lumpsum (Ultra) ->	292,88	20	982.070,48	984.171,00	986.271,52
Servers y SO->	27	5	78.549,02	92.410,61	104.423,99
Workstation (HP) ->	45	10	78.213,80	85.015,00	96.917,10
DDC (ACER) ->	72	10	73.320,00	78.000,00	92.040,00
Monitores 42->	71	10	96.360,00	109.500,00	122.640,00
Brackets for FIDS Monitors + Imstallation ->	65	10	44.500,00	50.000,00	58.000,00
LAN and VoIP - Materiales (Cisco / DOS) ->	194	5	974.763,16	975.713,08	983.734,30
LAN and VoIP - Servicios (Cisco / DOS) ->	444,63	15	190.144,05	208.949,50	231.933,95
Master Clock (Moba Time) ->	230	2	9.310,95	10.120,60	11.638,69
FIDS Labor JCI ->	188,88	10	323.921,36	344.597,19	406.624,68
FIDS FMT JCI ->	64	5	286.667,62	329.503,01	382.223,49
Security Systems: Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	287	20	975.718,44	983.393,53	995.238,30
Subcontract Installation ->	89	15	66.750,00	75.000,00	85.500,00
Security Labor JCI ->	104	20	190.000,00	200.000,00	226.000,00
Security FMT JCI ->	50	5	106.814,19	113.632,12	123.859,01

Public Address (Bosh) ->	90	10	353.400,00	380.000,00	399.000,00
Ingenieria PA (Bosh) ->	30	10	31.850,00	35.000,00	39.200,00
Notifier ->	120	5	246.297,60	256.560,00	289.912,80
Wattstopper ->	180	5	37.600,00	40.000,00	45.600,00
PA Labor JCI ->	110	10	105.411,39	114.577,60	127.181,14
PA FMT JCI ->	90	10	72428,72	79.592,00	94714,48
TOTALES:			5.324.090,78	5.545.735,24	5.902.653,45

La tabla 24 muestra la estimación más probable original con su valor de USD 5.545.735,24, mientras que el Gráfico 91 muestra los percentiles obtenidos en la simulación de Monte Carlo con 20 mil iteraciones, en este Gráfico se puede apreciar los niveles de riesgo existente, los mismos que son: Riesgo alto, riesgo medio y riesgo bajo.

La duración más probable de la tabla 24 es USD 5.545.735,24 y la probabilidad de lograr el objetivo es de 49,31% Tabla 28, y se encuentra en la zona de riesgo medio según Gráfico 91.

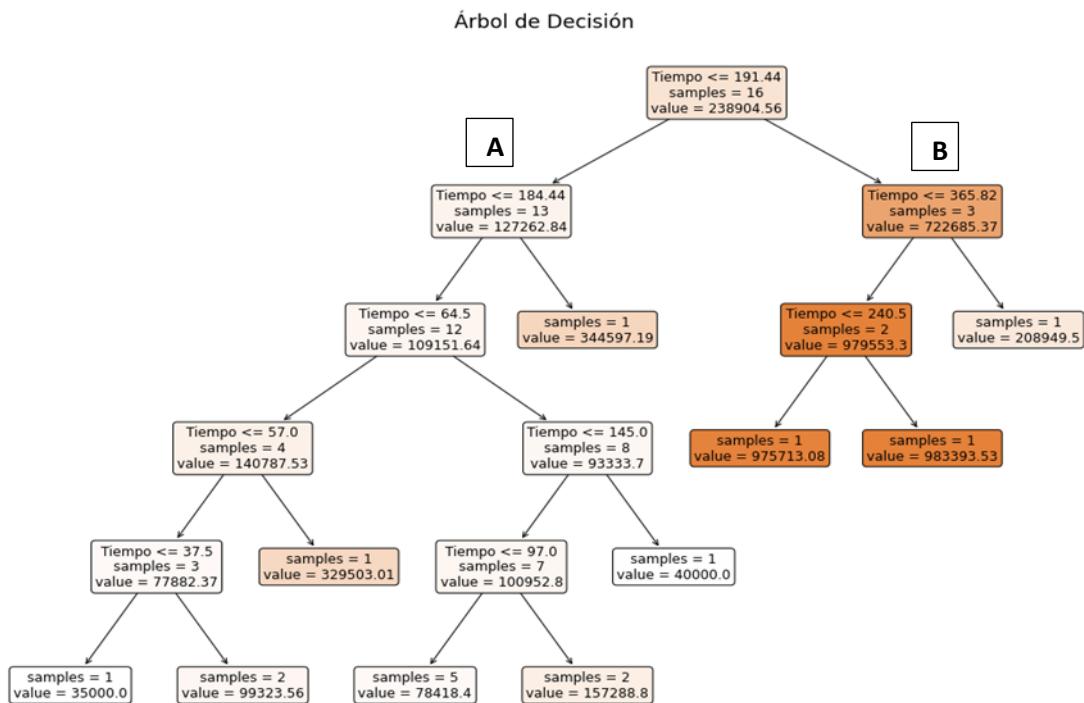
Se observa que la reserva de tiempo es de USD 5,85 según la tabla 28, lo que determina que el modelo es muy óptimo.

La desviación estándar del modelo es (+-) USD 358,18 según la tabla 28 indica que la dispersión en torno a la media es baja, lo que ratifica que existe un riesgo medio.

Se recomienda en base a los resultados obtenidos en el modelo, volver a estimar los costos de las actividades, queda a libertad del equipo ir hacia la izquierda o derecha de la media, ya dependerá si son adversos o arriesgados al riesgo (Osejo Francisco, 2023).

3.6.1.2.2 Árbol de Decisión

Gráfico 92 Arboles de Decisión con frecuencia - Presupuesto – Costo Autor: (Osejo Francisco, 2023)



El Gráfico 92 muestra dos ramales A y B, el modelo de machine learning árbol de decisión se divide en dos ramales A para actividades menores o iguales a 191,44 días y el ramal B para actividades mayores o iguales a 365,82 días (Osejo Francisco, 2023).

El modelo de árbol de decisión identifica con una precisión del 0,99 Tabla 26, la predicción de los costos – presupuesto y muestra al ramal B como el más idóneo, en este caso el de mayor riesgo, ya que, el ramal B tiene la predicción más alta en costo - presupuesto (Osejo Francisco, 2023).

Tabla 25 Actividades ramal B - árbol de decisión - presupuesto – costo Autor: (Osejo Francisco, 2023)

Actividad	Tiempo	Recursos	Mas_probable
FIDS/BIDS/RMS/AODB/CUTE - Lumpsum (Ultra) ->	292,88	20	984.171,00
LAN and VoIP - Materiales (Cisco / DOS) ->	194	5	975.713,08
Master Clock (Moba Time) ->	230	2	10.120,60
Security Systems: Access Control / CCTV / IC - Lumpsum (Visual Defence) ->	287	20	983.393,53

La tabla 25 muestra las actividades seleccionadas por el modelo de árbol de decisión, las cuales son consideradas con mayor riesgo. Para este caso, el número de actividades son pocas y puntuales, por lo que, se sugiere realizar el análisis de riesgos pertinente y reformulas planes de acción (Osejo Francisco, 2023).

CAPÍTULO IV: Análisis Comparativo de Modelos Aplicados

4 Análisis Comparativo de Modelos Aplicados

4.1 Simulación de Monte Carlo.

4.1.1 Simulación de Monte Carlo al Dataset de Cronograma – Tiempo

Se aplicó de forma independiente las Distribuciones Triangular y Beta-Pert.

En cada caso las iteraciones realizadas fueron de 1.000, 5.000, 10.000, 20.000

Tabla 26 Cuadro comparativo simulación monte carlo – dataset cronograma – tiempo.
Autor: (Osejo Francisco, 2023)

Distribución Triangular Iteraciones 1000, 5000, 10000, 20000		Distribución Beta-Pert Iteraciones 1000, 5000, 10000, 20000
Frecuencia	442,5	435
Media	236,921	236,134
Desviación estándar	3,259	2,872
Tiempo ejecución segundos	1,73	3,318
Reserva en número de días	1,921	1,134
Q0	225,130	226,114
Q25	234,738	234,163
Q50	236,910	236,118
Q75	239,116	238,057
Q95	242,214	240,861
Q100	249,075	247,382
Probabilidad de cumplir el cronograma objetivo, estimación Más Probable 235 días :	27,600%	34,715%
Probabilidad de cumplir el cronograma objetivo, estimación Optimista 207,28 días :	0,0%	0%
Probabilidad de cumplir el cronograma objetivo, estimación Pesimista 268,50 días :	100%	100%
Comprobación Q0,Q25, Q50, Q75, Q95,Q100		
Probabilidad Q0:	0%	0,00%
Probabilidad Q25:	25,033%	24,995%

Probabilidad Q50:	49,968%	49,993%
Probabilidad Q75:	74,970%	74,978%
Probabilidad Q95:	94,928%	94,965%
Probabilidad Q100:	99,965%	99,988%

La tabla 26 muestra los promedios obtenidos en cada iteración de 1.000, 5.000, 10.000 y 20.000 aplicados de manera independiente en el dataset de Cronograma – Tiempo. Los resultados obtenidos tanto en la Distribución Triangular, así como en la Distribución Beta – Pert se analizan a continuación:

- Frecuencia. - Muestra el valor máximo el en cual se agruparon los datos al momento de la ejecución del algoritmo. En este caso la Distribución Triangular tiene el mayor valor, pero al momento de este análisis no es un factor determinante.
- Media. - Corresponde al Q50 obtenido. En este caso la Distribución Beta Pert es más precisa.
- Desviación Standard. La Distribución Beta Pert es más precisa, describe un menor valor en la variación (+ -) de los rangos de desviación. Valor determinante.
- Tiempo ejecución en segundos. - Permite conocer el tiempo promedio que requirió la ejecución del algoritmo en cada iteración, valor referencial más no determinante.
- Reserva en días. - Es la diferencia entre la media obtenida Vs el valor más probable estimado inicialmente. Valor a tomar en cuenta como reserva o contingente. Podría estar inmerso en los rangos obtenidos en la desviación estándar, ya depende del profesional que realiza el análisis. En este caso la Distribución Beta – Pert es más precisa.
- Q0.- Es el valor mínimo en duración en días que puede iniciar el conjunto de actividades seleccionadas en el dataset. Valores menores a Q0 no aplica en el modelo. La interpretación del caso es duración en tiempo en este rango tiene cero probabilidades de lograr el objetivo y por lo tanto existen riesgos altos a tratar.

- Q50.- También llamado P50, es el valor medio obtenido del modelo. La interpretación es que existe la probabilidad del 50 % de conseguir el objetivo. Identifica que existe riesgos medios a tomar en cuenta.
- Q100.- Se interpreta como la duración máxima permitida en el modelo, cualquier valor mayor se interpreta como erróneo y la presencia de riesgo alto. También indica que existe el 100% probabilidad de lograr el objetivo y obviamente riesgo bajo.
- Probabilidad de cumplir la estimación de duración más probable original. - En este caso la Distribución Beta-Pert tiene un resultado más alto y preciso, a pesar de aquello el **34,715%** sigue siendo inquietante y de alto riesgo, por lo que, se recomienda revisar las estimaciones originales en base a las obtenidas y seguir iterando hasta tener el modelo más óptimo.
- Probabilidad de cumplir la estimación de duración optimista original. - Q0 es mayor a la estimación optimista original, por lo que, tiene cero posibilidades de conseguir el objetivo y por ende el riesgo es muy alto. Se recomienda retornar a la mesa de trabajo y obtener mejores estimaciones.
- Probabilidad de cumplir la estimación de duración pesimista original. La probabilidad es mayor al 100%, ya que, Q100 es menor a la duración pesimista. Se comprende que la duración pesimista es el valor máximo probable. Por ende, existe incertidumbre en el modelo.
- Comprobación Q0, Q25, Q50, Q75, Q95, Q100.- Se procedió a verificar la probabilidad de los percentiles obtenidos, la diferencia existente se debe a los decimales tomados al momento de realizar el cálculo pertinente.

En conclusión, La distribución Beta-Pert es la más precisa y se debe a que no es un promedio ponderado como lo es la Distribución Triangular (Osejo Francisco, 2023).

4.1.2 Simulación de Montecarlo al Dataset de Presupuesto – Costo

Se aplicó de forma independiente las Distribuciones Triangular y Beta-Pert.

En cada caso las iteraciones realizadas fueron de 1.000, 5.000, 10.000, 20.000

Tabla 27 Cuadro comparativo simulación monte carlo – dataset presupuesto – Costo.
Autor: (Osejo Francisco, 2023)

Distribución Triangular Iteraciones 1000, 5000, 10000, 20000		Distribución Beta-Pert Iteraciones 1000, 5000, 10000, 20000
Frecuencia	447,5	463,75
Media	5.590.654,14	5.545.741,09
Desviación estándar	34.319,90	358,18
Tiempo ejecución segundos	1,7875	2,95
Reserva en número de días	44.918,90	5,85
Q0	5.474.977,35	5.544.413,17
Q25	5.566.642,41	5.545.500,87
Q50	5.590.320,32	5.545.740,44
Q75	5.613.921,23	5.545.984,17
Q95	5.648.138,46	5.546.327,92
Q100	5.720.388,66	5.547.160,30
Probabilidad de cumplir el costo objetivo, estimación Más Probable 5.545.735,24 dólares :	9,53%	49,31%
Probabilidad de cumplir el costo objetivo, estimación Optimista 5.324.090,777 dólares :	0,0%	0,0%
Probabilidad de cumplir el costo objetivo, estimación Pesimista 5,902,653,451 dólares :	100%	100%
Comprobación Q0,Q25, Q50, Q75, Q95,Q100		
Probabilidad Q0:	0%	0%
Probabilidad Q25:	25%	25%
Probabilidad Q50:	50%	50%
Probabilidad Q75:	75%	75%
Probabilidad Q95:	95%	95%
Probabilidad Q100:	100%	100%

La tabla 27 muestra los promedios obtenidos en cada iteración de 1.000, 5.000, 10.000 y 20.000 aplicados de manera independiente en el dataset de Presupuesto – Costo. Los

resultados obtenidos tanto en la Distribución Triangular, así como en la Distribución Beta – Pert se analizan a continuación:

- Frecuencia. - Muestra el valor máximo el en cual se agruparon los datos al momento de la ejecución del algoritmo. En este caso la Distribución Triangular tiene el mayor valor, pero al momento de este análisis no es un factor determinante.
- Media. - Corresponde al Q50 obtenido. En este caso la Distribución Beta Pert es más precisa.
- Desviación Standard. La Distribución Beta Pert es más precisa, describe un menor valor en la variación (+ -) de los rangos de desviación. Valor determinante.
- Tiempo ejecución en segundos. - Permite conocer el tiempo promedio que requirió la ejecución del algoritmo en cada iteración, valor referencial más no determinante.
- Reserva en dólares. - Es la diferencia entre la media obtenida Vs el valor más probable estimado inicialmente. Valor a tomar en cuenta como reserva o contingente. Podría estar inmerso en los rangos obtenidos en la desviación estándar, ya depende del profesional que realiza el análisis. En este caso la Distribución Beta – Pert es más precisa.
- Q0.- Es el valor mínimo en presupuesto que puede costar el conjunto de actividades seleccionadas en el dataset. Costos menores a Q0 no aplica en el modelo. La interpretación del caso es presupuesto en dólares, en este rango tiene cero probabilidades de lograr el objetivo y por lo tanto existen riesgos altos a tratar.
- Q50.- También llamado P50, es el valor medio obtenido del modelo. La interpretación es que existe la probabilidad del 50 % de conseguir el objetivo. Identifica que existe riesgos medios a tomar en cuenta.
- Q100.- Se interpreta como presupuesto máximo permitido en el modelo, cualquier valor mayor se interpreta como erróneo y la presencia de riesgo alto. También indica que existe el 100% probabilidad de lograr el objetivo y obviamente riesgo bajo.

- Probabilidad de cumplir la estimación de duración más probable original. - En este caso la Distribución Beta-Pert tiene un resultado más alto y preciso, a pesar de aquello el **49,31%** sigue siendo inquietante y de alto riesgo, por lo que, se recomienda revisar las estimaciones originales en base a las obtenidas y seguir iterando hasta tener el modelo más óptimo.
- Probabilidad de cumplir la estimación del presupuesto optimista original. - Q0 es mayor a la estimación optimista original, por lo que, tiene cero posibilidades de conseguir el objetivo y por ende el riesgo es muy alto. Se recomienda retornar a la mesa de trabajo y obtener mejores estimaciones.
- Probabilidad de cumplir la estimación del presupuesto pesimista original. La probabilidad es mayor al 100%, ya que, Q100 es menor a la duración pesimista. Se comprende que la duración pesimista es el valor máximo probable. Por ende, existe incertidumbre en el modelo.
- Comprobación Q0, Q25, Q50, Q75, Q95, Q100.- Se procedió a verificar la probabilidad de los percentiles obtenidos, la diferencia existente se debe a los decimales tomados al momento de realizar el cálculo pertinente.

En conclusión, La distribución Beta-Pert es la más precisa y se debe a que no es un promedio ponderado como lo es la Distribución Triangular (Osejo Francisco, 2023).

4.2 Resumen del análisis comparativo de los modelos aplicados

Tabla 28 Tabla Resumen análisis comparativo de modelos aplicados. Autor: (Osejo Francisco, 2023)

Algoritmo	Cronograma Tiempo	Presupuesto Costo	Aplica para Análisis de Riesgo Cuantitativo	Resultados	Observación
Simulación Montecarlo	X	X	Sí	Percentiles Q0, Q25, Q50, Q75, Q100 Óptimos	Los resultados obtenidos permiten comparar los valores estimados iniciales: optimista, más probable y pesimista con los datos obtenidos en la simulación de Montecarlo, permite identificar los riesgos positivos como negativos.
Regresión Lineal Simple	N/A	X	NO	Coeficiente de determinación R^2 = 0.24177	Valor obtenido del Coeficiente de determinación R^2 del modelo por debajo del 0,50. Determina que Regresión Lineal Simple no aplica para este Dataset, ya que, existe alta dispersión en los datos.
Regresión Lineal Múltiple	N/A	X	NO	Coeficiente de determinación R^2 = 0.2236	Valor obtenido del Coeficiente de determinación R^2 del modelo por debajo del 0,50. Determina que Regresión Lineal Simple no aplica para este Dataset, ya que, existe alta dispersión en los datos.
Árbol de Decisión	N/A	X	SI	Precisión del modelo 0,99578	Precisión del modelo es muy alta cerca a 1, adicionalmente la predicción es muy buena y permite a través de sus ramales identificar perfectamente los riesgos positivos y negativos.

Simulación Montecarlo. - Este algoritmo demostró ser muy preciso y didáctico al momento de realizar las iteraciones de 1,000, 5,000, 10,000, 20,000. En cada iteración fue afinando los datos demostrando mayor precisión. Excelente al momento de realizar el análisis de los percentiles Q0, Q25, Q75, Q100 (Osejo Francisco, 2023).

Se aplicó las Distribuciones Triangular y Beta – Pert para ambos datasets tanto para Cronograma, así como para el Presupuesto. En el capítulo 4.1 se muestra los datos levantados y su análisis comparativo (Osejo Francisco, 2023).

Regresión Lineal Simple. - No aplicó para el dataset de Cronograma – tiempo ya que la variable dependiente tiempo tenía una alta relación con las variables independientes,

tales como: tiempo optimista, más probable y pesimista, lo que generaría multicolinealidad (Osejo Francisco, 2023).

En cambio, el dataset Presupuesto – Costo si se pudo aplicar el algoritmo de regresión lineal simple siendo presupuesto la variable dependiente a predecir en base al comportamiento de la variable independiente Tiempo. El coeficiente de Determinación R² obtenido es: 0,25177 siendo una precisión muy baja, por lo que, se determina que el modelo aplicando regresión lineal simple no es óptimo para el dataset de Presupuesto – Costo (Osejo Francisco, 2023).

Regresión Lineal Múltiple. - No aplicó para el dataset de Cronograma – tiempo ya que la variable dependiente tiempo tenía una alta relación con las variables independientes, tales como: tiempo optimista, más probable y pesimista, lo que generaría multicolinealidad (Osejo Francisco, 2023).

En cambio, el dataset Presupuesto – Costo si se pudo aplicar el algoritmo de regresión lineal simple siendo presupuesto la variable dependiente a predecir en base al comportamiento de las variables independientes: Tiempo y Recursos. El coeficiente de Determinación R² obtenido es: 0,2236 siendo una precisión muy baja, por lo que, se determina que el modelo aplicando regresión lineal múltiple no es óptimo para el dataset de Presupuesto – Costo (Osejo Francisco, 2023).

Arboles de Decisiones.- No aplicó para el dataset de Cronograma – tiempo ya que la variable dependiente tiempo tenía una alta relación con las variables independientes, tales como: tiempo optimista, más probable y pesimista, lo que generaría multicolinealidad (Osejo Francisco, 2023).

Se aplicó al dataset de presupuesto – costo y se obtuvo una precisión del modelo de 0,99578 siendo un excelente valor, se observó cómo cada ramal nos ayudaba a ver el comportamiento de las frecuencias y su predicción (Osejo Francisco, 2023).

Como resultado se logró identificar el ramal con mayor riesgo e impacto, junto con su predicción (Osejo Francisco, 2023).

CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES

5 CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

1. Los algoritmos de Machine Learning más relevantes que se aplican a la gestión de riesgos cuantitativos en proyectos predictivos adaptativos analizados en este trabajo de titulación, fueron seleccionados de fuentes de investigación como libros y publicaciones especializadas en el ámbito de la Gestión de Riesgos. Los algoritmos seleccionados son:
 - a) Monte Carlo
 - b) Árboles de decisión
 - c) Análisis de regresión
2. Mediante la aplicación de los algoritmos de Machine Learning, utilizados en este trabajo de titulación, se ha podido identificar las diferencias entre riesgos cualitativos vs riesgos cuantitativos.
3. La información del proyecto implementado en el nuevo aeropuerto de Quito, sirvió de base para la conformación de los datasets, específicamente se utilizó la información basada en cronograma de actividades con duraciones – tiempos y otro con presupuesto – costos relacionados al mismo proyecto, siendo una pequeña muestra de la automatización del nuevo aeropuerto Internacional de Quito. Además, los datasets permitieron recrear el análisis comparativo de la aplicación de algoritmos de machine learning en la gestión de riesgos cuantitativos, determinando que no todos los algoritmos en cuestión pudieron ser aplicados efectivamente ni obtener el resultado esperado, lo que, conlleva a concluir que los modelos aplicados no son los apropiados para los datasets seleccionados.
4. Se concluye que la Metodología CRISP DM permitió estructurar de forma ordenada el desarrollo de este trabajo de titulación, tanto en el tratamiento de los diferentes datasets así como en modelado y análisis de resultados obtenidos.
5. Se logró realizar el análisis comparativo y establecer las conclusiones de la aplicación de algoritmos de machine learning seleccionados en la aplicación de

la gestión de riesgos cuantitativos en proyectos predictivos y adaptativos, cuyos resultados permiten concluir que la simulación de Monte Carlo es la mejor opción para el análisis de riesgo cuantitativo para los datasets seleccionados, con el modelo de regresión simple y múltiple no se obtuvo la precisión deseada y finalmente el modelo de Árboles de Decisión presentó una alta precisión además su presentación gráfica permite visualizar de manera eficiente el comportamiento de la data, otorgando mejor capacidad de análisis.

6. El trabajo realizado ha sido de gran valor, ya que ha permitido afinar conceptos que al poner en práctica demostró la importancia de la aplicación de Machine Learning.

5.2 Recomendaciones

1. Se recomienda tomar en cuenta los resultados obtenidos con los modelos de Machine Learning de regresión simple y múltiple para aplicar el mismo principio con otros datasets que cuenten con mayor cantidad de variables independientes, lo que permitirá determinar la verdadera aplicación y efectividad del modelo en cuestión.
2. Se recomienda ampliar este estudio realizado aplicando otros algoritmos de Machine Learning de predicción con el afán de instaurar una práctica común.
3. Es importante tomar en cuenta los conceptos de Inteligencia Artificial como la Lógica Difusa, principio fundamental que nos permite ampliar el espectro de comparación al permitir ponderar los riesgos con mayor objetividad, por lo que, se recomienda ampliar el estudio con el afán de lograr obtener mayores aplicaciones en este ámbito.
4. Se recomienda prepararse, investigar, profundizar más con la finalidad de aplicar estos conocimientos en la vida diaria e instaurar la adopción de IA como una forma de crecer profesionalmente, así como económicoamente.

BIBLIOGRAFÍA

- Ernst & Young, 2009. Encuesta de EY de principales proyectos de capital de construcción de compañías Fortune 500.
- MacGraw-Hill, 2011. Mitigation of Risk in Construction.
- Project Management Institute, 2021. Guía de los Fundamentos para la Dirección de Proyectos – Guía PMBOK Séptima Edición.
- Project Management Institute, 2017. Guía de los Fundamentos para la Dirección de Proyectos – Guía PMBOK Sexta Edición.
- Liliana Buchtik, 2015. Secretos para Dominar La Gestión de Riesgos en Proyectos.
- Liliana Buchtik, 2016. Secretos para Dominar La Gestión de Portafolios de Programas y Proyectos.
- Liliana Buchtik, 2023. Secretos para Dominar La Gestión de Riesgos Emergentes.
- Project Management Institute, 2011. Practice Standard for Project Estimating.
- Project Management Institute, 2009. Practice Standard for Project Risk Management.
- Project Management Institute, 2010. Pulse of the Profession Research.
- ISO 31000, 2009. Norma Internacional para la Gestión del Riesgo.
- Nassim, 2010. The Black Swan: The Impact of the Highly Improbable.
- Lógica Difusa, 2017. Lógica Difusa.
- <http://usslogicadifusa.blogspot.com/2017/09/que-es-lalogica-difusa-basicamente-la.html>
- Flach, P. (2012). MACHINE LEARNING: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press.
- Alpaydin, E., (2014). Introducción a Machine Learning. Londres: MIT Press.

McKinsey&Company, (2018). Notes From the AI Frontier Insights from Hundreds of Use Cases.

<https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning#part4>

Brian Boucheron & Lisa Tagliaferri, 2022. Python Machine Learning Projects.

Carlos N. Bouza, 2018. Modelos de regresión y Sus Aplicaciones.

IBM, 2022. ¿Qué es la simulación Monte Carlo?.

<https://www.ibm.com/mx-es/topics/monte-carlo-simulation>

SciPy.org, 2023 Documentación SciPy.

<https://docs.scipy.org/doc/scipy/>

Quiport, 2017. Concesión Del Aeropuerto Internacional de Quito.

<https://www.quiport.com/es/naiq.html>