

Tabla de Contenido

1. Introducción	1
2. Desarrollo	4
2.1. Antecedentes de Estudio	4
2.2. Metodología	7
2.3. Limitaciones	8
2.4. Datos	9
2.4.1. Descripción de las Bases de Datos	9
2.4.1.1. DEMRE	9
2.4.1.2. FCFM	10
2.4.1.3. MINEDUC	10
2.4.2. Cruce de Bases de Datos	10
2.4.3. Creación de variables	12
2.4.3.1. Creación de la variable a pronosticar	17
2.4.4. Preprocesamiento	17
2.4.5. Variables disponibles para ser ocupadas en los diversos modelos	18
2.4.6. Desbalance de clases	22
2.5. Modelos	23
2.5.1. Support Vector Machine (SVM)	24
2.5.2. Árbol de Clasificación (DT)	25
2.5.3. Random Forest (RF)	25
2.5.4. Red Neuronal (RN)	25
2.5.5. K vecinos más cercanos (KNN)	26
2.5.6. Regresión Logística (Logit)	26

2.5.7.	Ajustes Generales	27
3.	Resultados	28
3.1.	Predicciones con datos previos al ingreso a Plan Común	28
3.1.1.	Variables importantes en la predicción	35
3.1.1.1.	Análisis RN	35
3.1.1.2.	Análisis Logit	38
3.2.	Predicciones con datos de rendimiento parcial de Plan Común	39
3.2.1.	Modelos considerando tasa de aprobación 1º Semestre	40
3.2.1.1.	Variables importantes para los modelos considerando la tasa de aprobación del 1º Semestre	46
3.2.1.1.1	Análisis RN	47
3.2.1.1.2	Análisis Logit	49
3.2.1.1.3	Análisis DT	51
3.2.2.	Modelos considerando tasa de aprobación 1º y 2º Semestre	52
3.2.2.1.	Variables importantes para los modelos considerando la tasa de aprobación del 1º y 2º Semestre.	58
3.2.2.1.1	Análisis Logit	58
3.2.2.1.2	Análisis RF	60
3.2.3.	Evolución de los modelos	61
3.2.4.	Discusión	63
4.	Conclusiones	66
	Bibliografía	68
	Anexos	74
A.	Información descriptiva	74
B.	Información descriptiva de las Tasas de aprobación	78
C.	Resultados de los modelos	79
C.1.	KNN	79
C.2.	SVM	80
C.3.	DT	80

C.4.	RF	81
C.5.	RN	82
C.6.	Logit	83
D.	Resultados de los modelos, considerando la Tasa de Aprobación del 1º Semestre	83
D.1.	KNN, considerando la Tasa de Aprobación del 1º Semestre	83
D.2.	SVM, considerando la Tasa de Aprobación del 1º Semestre	84
D.3.	DT, considerando la Tasa de Aprobación del 1º Semestre	85
D.4.	RF, considerando la Tasa de Aprobación del 1º Semestre	86
D.5.	RN, considerando la Tasa de Aprobación del 1º Semestre	87
D.6.	Logit, considerando la Tasa de Aprobación del 1º Semestre	88
E.	Resultados de los modelos, considerando las Tasas de aprobación del 1º y 2º Semestre	89
E.1.	KNN, considerando las Tasas de aprobación del 1º y 2º Semestre . . .	89
E.2.	SVM, considerando las Tasas de aprobación del 1º y 2º Semestre . . .	90
E.3.	DT, considerando las Tasas de aprobación del 1º y 2º Semestre	91
E.4.	RF, considerando las Tasas de aprobación del 1º y 2º Semestre	92
E.5.	RN, considerando las Tasas de aprobación del 1º y 2º Semestre	93
E.6.	Logit, considerando las Tasas de aprobación del 1º y 2º Semestre . . .	94