



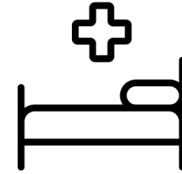
PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Predicción del tiempo de permanencia de pacientes en un hospital



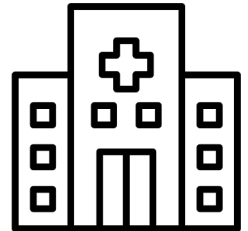
1. Introducción

- Asignación de recursos en hospitales
- Mejorar la atención de pacientes en cantidad y calidad
- Recursos finitos



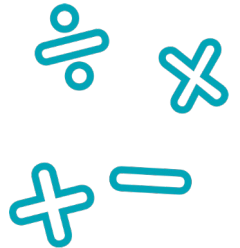
1.1 ¿Cuál es el principal problema asociado?

- Asignación eficiente de camas para pacientes de urgencia o en espera
- Resolverlo estimando los tiempos de permanencia de pacientes en un hospital. En otras palabras, **estimando el LOS (length of stay) de los pacientes**



1.2 Complejidades del problema

- Diversos factores influyen en el LOS
- Variación de LOS según centro de salud
- La asignación eficiente de camas está sujeta al tiempo de permanencia
- Pacientes con alto LOS implica una disminución en la capacidad del hospital



2.1 Descripción Datos

Procedures

Consta de 4 variables principales

- **Case:** ID paciente
- **Date:** Fecha procedimiento
- **Release:** Fecha de alta
- **Procedure:** Código

Diagnosis

Consta de 4 variables principales

- **Case :** ID paciente
- **PrincSec :** Tipo de diagnóstico
- **Seq :** Número diagnostico principal asociado
- **Diagnosis :** Código



2.2 Descripción del problema

Alcances

- Reducción tiempos de espera
- Descongestión de sistemas de salud
- Planificación interna más eficiente
- Manejo eficiente de insumos médicos



3. Metodologías

Utilizadas

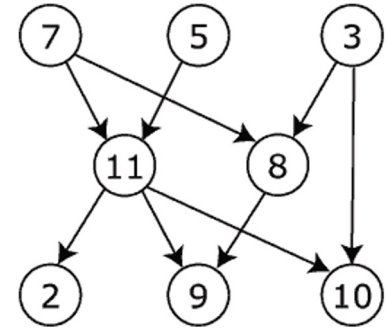
- Random Forest
 - K-Means **(12 intervalos)**
- Naive Bayes
 - K-Means **(12 intervalos)**
- Decision Tree
- Support Vector Machine
- Multilayer Perceptron



3. Metodologías

Redes bayesianas

- Grafo acíclico dirigido
- **Aristas** representan **dependencias condicionales** y los **nodos** son **estados**
- Enfoque ***Maximum a Posteriori (MAP)***



3. Metodologías

Naive Bayes

- **Independencia** entre variables **predictoras**
- Se puede **entrenar de manera eficiente** en un entorno de **aprendizaje supervisado**
- Se requieren **pocos datos de entrenamiento** para **estimar parámetros**



4. Procesamiento de Datos

Complejidad Inicial

- Pacientes con muchos procedimientos que tenían **LOS más pequeños que su LOS total (distinta entrada, misma salida)**.
- Esto se debe a que fueron procedimientos que se fueron realizando **a medida que el paciente se encontraba en el hospital**

4. Procesamiento de Datos

Complejidad Inicial

- Dado el sistema de información del hospital, **no se sabe la contribución real** de cada uno de estos procedimientos en la estadía del paciente, solo se sabe que el conjunto de procedimiento gatilla un largo de estadía
- **Se decidió ajustar los LOS por paciente.**

4. Procesamiento de Datos

Complejidad Inicial: Ejemplo Paciente 14158735

Case	Procedure	Date	Release	LOS
14158735	3E04305	01-08-2017	14-01-2018	166
14158735	3E0G76Z	26-11-2017	14-01-2018	49
14158735	5A1955Z	25-11-2017	14-01-2018	50
14158735	5A1955Z	17-12-2017	14-01-2018	28
14158735	8E0ZXY6	29-11-2017	14-01-2018	46
14158735	DB021ZZ	24-07-2017	14-01-2018	174

4. Procesamiento de Datos

Complejidad Inicial: Ejemplo Paciente 14158735

Case	Procedure	Date	Release	LOS	LOS Final
14158735	3E04305	01-08-2017	14-01-2018	186	186
14158735	3E0G76Z	26-11-2017	14-01-2018	49	186
14158735	5A1955Z	25-11-2017	14-01-2018	50	186
14158735	5A1955Z	17-12-2017	14-01-2018	28	186
14158735	8E0ZXY6	29-11-2017	14-01-2018	46	186
14158735	DB021ZZ	24-07-2017	14-01-2018	174	186

4. Procesamiento de Datos

Base de datos : Procedimientos

- Se revisó la cantidad de datos por procedimiento.
- Se agruparon los procedimientos por el primer carácter.

Procedimiento	0	1	2	3	4	5	6	8	B	C	D	F	G	H	X
Cantidad de datos	16275	1514	161	3048	2194	1165	78	94	1795	36	38	117	9	43	1

4. Procesamiento de Datos

Base de datos : Procedimientos

- Se volvió a agrupar el procedimiento 0 en 6 categorías por la desviación y cantidad de datos.
- Se agruparon el grupo de procedimientos de un paciente en una misma celda.
- Asignación de un orden según frecuencias de los distintos procedimientos
- Se hizo un *join()* con el excel de diagnósticos.

	Cantidad Datos	Intervalo LOS
0T	4247	[0,1]
0U	2278]1, 2]
0V	3475]2, 5]
0W	2061]5, 10]
0Y	2153]10, 25]
0Z	2061]25,]

4. Procesamiento de Datos

Base de datos : Diagnósticos

- Se decidió trabajar con los procedimientos según los diagnósticos principales de los pacientes.
- Se categorizaron por el primer carácter.
- Se agruparon como una tupla por paciente.

4. Procesamiento de Datos

Base de datos : Procedimientos y Diagnósticos

	Case	Diagnosis	Procedure	LOS
10	14198020	(I, K)	0Z-3	224
11	14228157	(E,)	0V	3
12	14229907	(C, I, U)	0Z	133
13	14234686	(E,)	0W	7
14	14236115	(M,)	0U	2
15	14240255	(A, I, U, Z)	0Z-3-4-5-B	159
16	14254687	(L,)	0V	4
17	14255134	(L,)	0T	1
18	14263128	(N,)	0T	1
19	14266798	(O,)	1	3
20	14266806	(O,)	0W-1-3-4	6
21	14266816	(O,)	0V-1	3
22	14266824	(O,)	0V-1-3-4	3
23	14267839	(C, U)	0Z-3	156

	Case	Diagnosis	Procedure	LOS
24	14275725	(S,)	0T	1
25	14281431	(D, Z)	0Z	115
26	14282974	(L,)	0V	3
27	14283280	(O,)	0V-1-3	4
28	14290400	(O,)	1-3-4	3
29	14296420	(M,)	0V	5
30	14300013	(O,)	0V-1	3
31	14302745	(M,)	0T	1
32	14303159	(L,)	0V	3
33	14308438	(O,)	0V-1-3-4	5
34	14308689	(O,)	0U-1-3-4	2
35	14308738	(O,)	1-3-4	2
36	14308831	(O,)	1-3-4	2
37	14308845	(O,)	1-3	3

4. Procesamiento de Datos

Base de datos: 11951 datos de pacientes

Procedimiento	Cantidad
0T	2645 (15.3%)
0U	1481 (8.55%)
0V	2370 (13.7%)
0W	1122 (6.48%)
0Y	924 (5.33%)
0Z	576 (3.32%)
1	1060 (6.12%)

Procedimiento	Cantidad
2	53 (0.30%)
3	2182 (12.6%)
4	2157 (12.45%)
5	953 (5.5%)
6	72 (0.42%)
8	94 (0.54%)
B	1410 (8.14%)

Procedimiento	Cantidad
C	34 (0.19%)
D	33 (0.19%)
F	113 (0.65%)
G	1 (0.006%)
H	42 (0.24%)
X	1 (0.006%)

4. Procesamiento de Datos

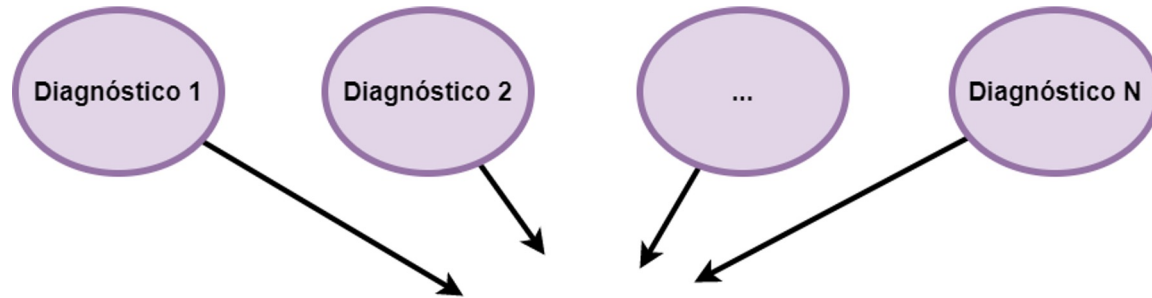
Base de datos: 11526 datos de pacientes

Procedimiento	Cantidad
0T	2634 (16.1%)
0U	1477 (9.04%)
0V	2334 (14.28%)
0W	1081 (6.61%)
0Y	851 (5.21%)
0Z	444 (2.72%)
1	1059 (6.48%)

Procedimiento	Cantidad
3	2034 (12.45%)
4	2149 (13.15%)
5	889 (5.44%)
B	1388 (8.5%)

5. Formulación Modelo

Diagnóstico : Probabilidades a priori



5. Formulación Modelo

Diagnóstico : Probabilidades a priori

Diagnóstico	Probabilidad
A	0.0158
B	0.0031
C	0.0616
D	0.0332
E	0.0475
F	0.0039

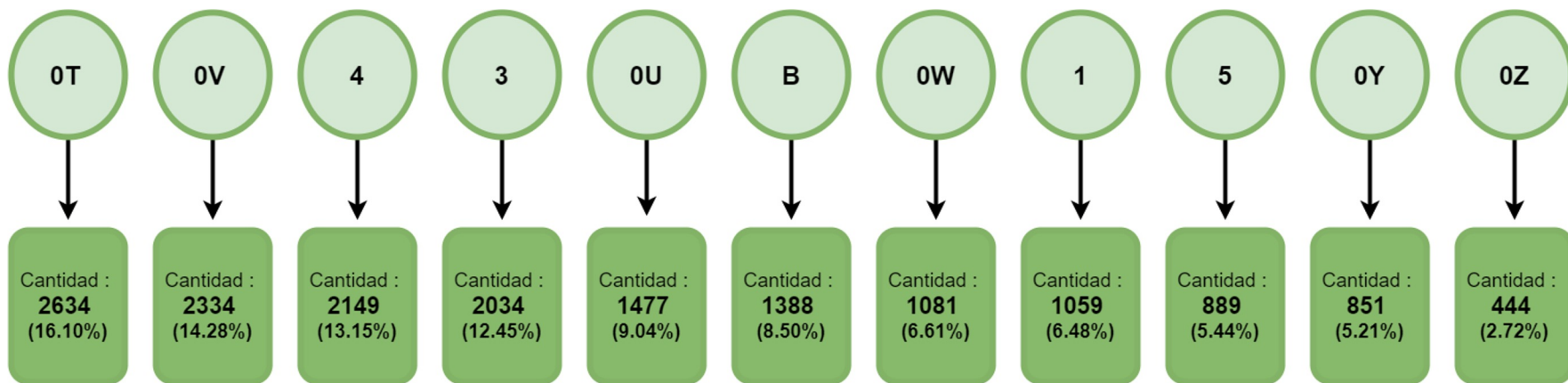
Diagnóstico	Probabilidad
G	0.0874
H	0.0056
I	0.1491
J	0.0648
K	0.1204
L	0.0108

Diagnóstico	Probabilidad
M	0.1091
N	0.0559
O	0.0979
P	0.0043
Q	0.0049
R	0.0281

Diagnóstico	Probabilidad
S	0.0704
T	0.0329
U	0.2855
Z	0.0938

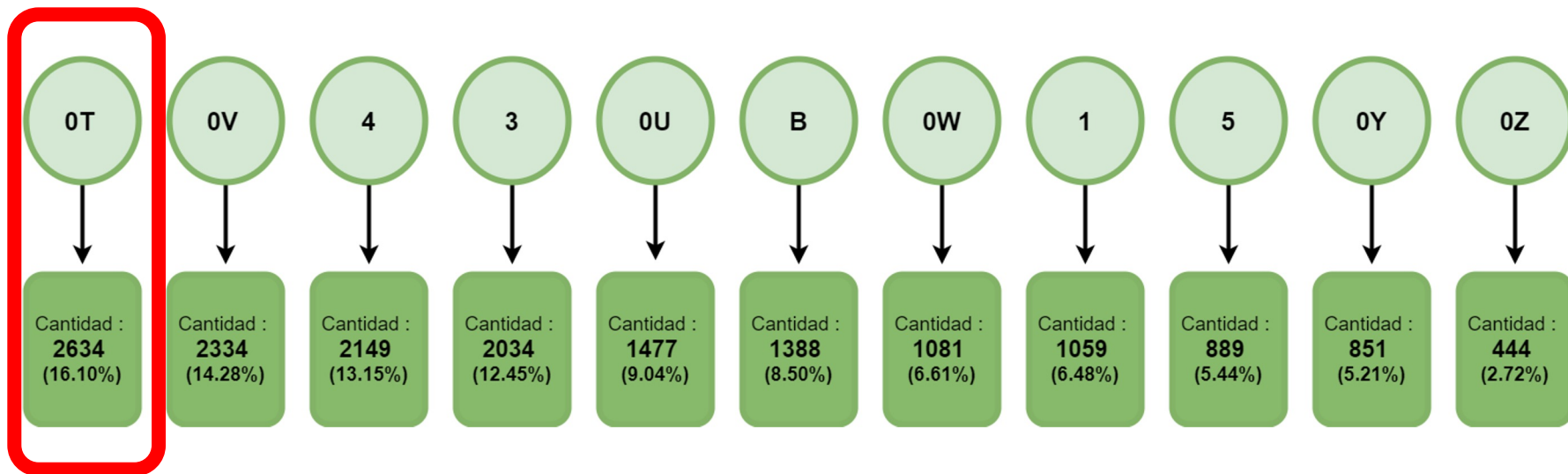
5. Formulación Modelo

Procedimiento : Orden por mayor ocurrencia



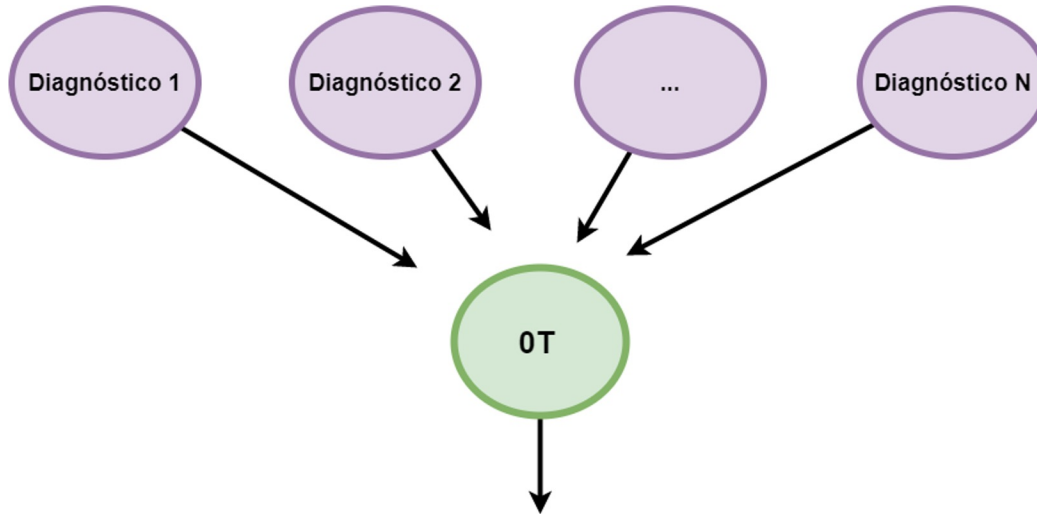
5. Formulación Modelo

Procedimiento : Orden por mayor ocurrencia



5. Formulación Modelo

Procedimiento : Orden por mayor ocurrencia



5. Formulación Modelo

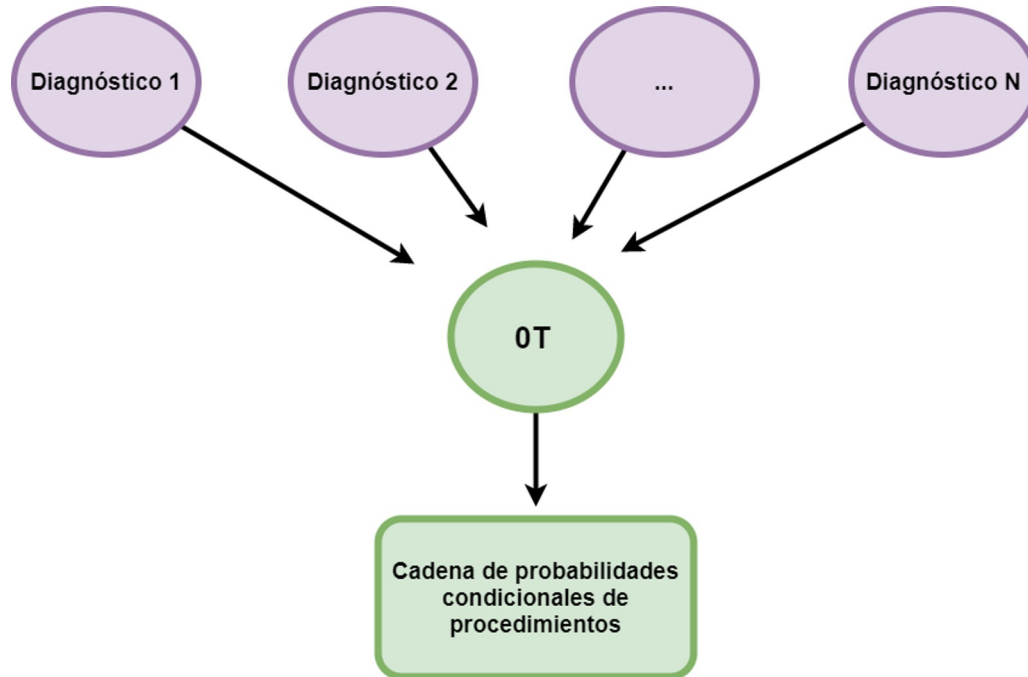
Procedimiento : Prob. condicionadas por nodo padre

Procedimiento (P =)	P Padre	\overline{P} Padre	P $\overline{\text{Padre}}$	\overline{P} $\overline{\text{Padre}}$
0V	0.000	1.000	0.262	0.738
4	0.081	0.919	0.213	0.787
3	0.119	0.881	0.188	0.812
0U	0.046	0.954	0.147	0.853
B	0.059	0.941	0.129	0.871

Procedimiento (P =)	P Padre	\overline{P} Padre	P $\overline{\text{Padre}}$	\overline{P} $\overline{\text{Padre}}$
0W	0.062	0.938	0.096	0.904
1	0.015	0.985	0.099	0.901
5	0.000	1.000	0.085	0.915
0Y	0.161	0.839	0.067	0.933
0Z	0.000	1.000	0.041	0.959

5. Formulación Modelo

Procedimiento : Prob. condicionadas por nodo padre

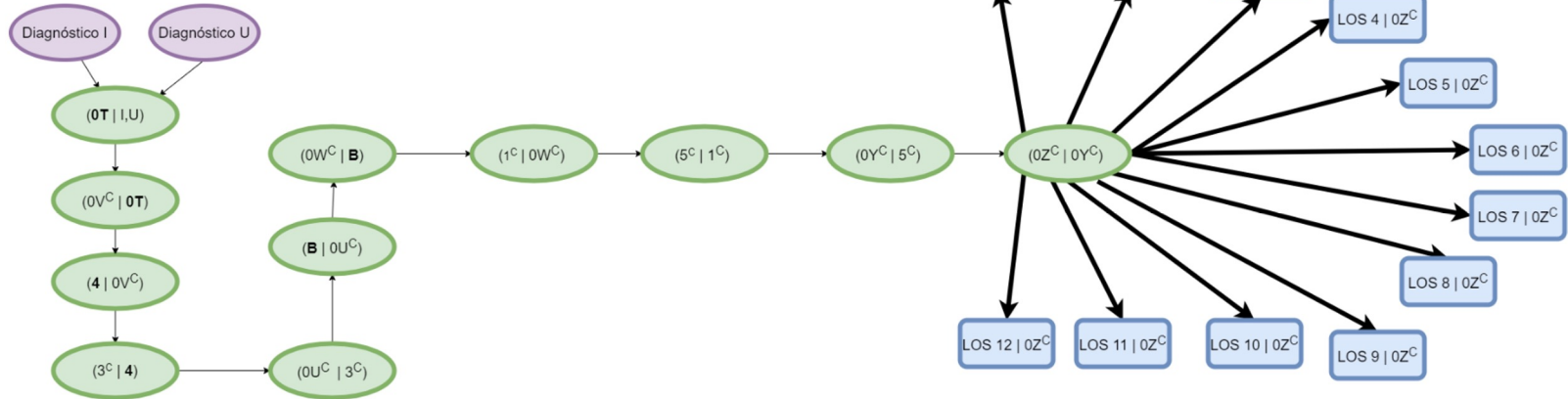


5.1 Ejemplo Modelamiento: Paciente 14574357

CASE: 14574357
DIAGNOSIS: I-U
PROCEDURE: OT-4-B

Orden Frecuencia Procedimientos :
OT > OV > 4 > 3 > OU > B > OW > 1 > 5 > OY > OZ

Case: 14574357



$$P(A^C) = P(\bar{A}) = 1 - P(A)$$

5.1 Ejemplo Resultados: Paciente 14574357

Intervalo LOS real (0.0, 5.053)
LOS real 3 días

Intervalo LOS	Valor Intervalo	Probabilidad
LOS 1 OZ ^C	(0.0, 5.053)	4.968e-05
LOS 2 OZ ^C	(5.053, 12.141)	8.357e-06
LOS 3 OZ ^C	(12.141, 21.285)	3.574e-06
LOS 4 OZ ^C	(21.285, 32.089)	9.325e-07
LOS 5 OZ ^C	(32.089, 44.246)	5.651e-08
LOS 6 OZ ^C	(44.246, 58.306)	2.119e-08

Intervalo LOS	Valor Intervalo	Probabilidad
LOS 7 OZ ^C	(58.306, 75.144)	2.119e-08
LOS 8 OZ ^C	(75.144, 93.408)	0
LOS 9 OZ ^C	(93.408, 110.25)	0
LOS 10 OZ ^C	(110.25, 125.75)	7.064e-09
LOS 11 OZ ^C	(12.75, 145.5)	0
LOS 12 OZ ^C	(145.5, 159)	0

6. Resultados y Discusión

Porcentajes obtenidos luego de realizar 10-Fold Cross Validation

Modelo Predictivo	Accuracy
Naive Bayes	77.75%
Random Forest	88.82 %
Decision Tree	67.49 %
Support Vector Machine	85.83 %
Multilayer Perceptron	85.76%

6. Resultados y Discusión

Métodos de Comparación de los Clasificadores

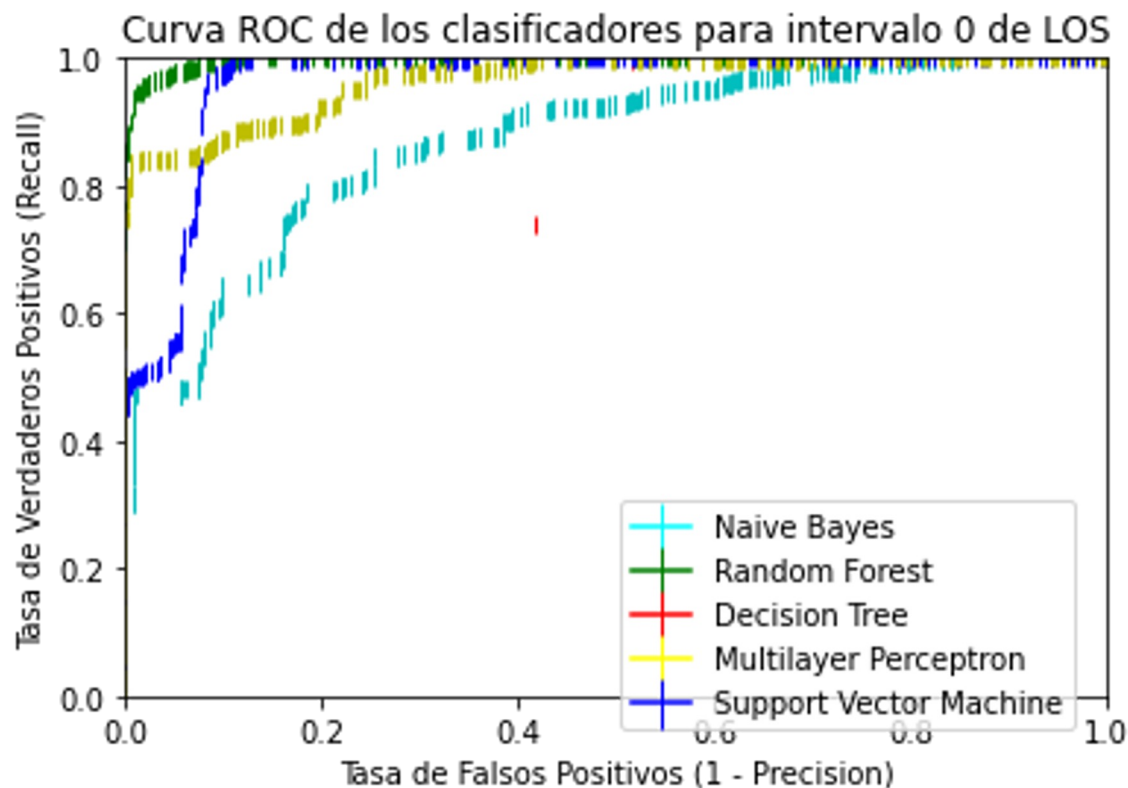
- **Curva ROC de los clasificadores** → Sensibilidad vs Especificidad
- **Accuracy** → Fracción de veces que el modelo acierta.
- **Weighted Precision** → Fracción de veces que no se comete error tipo II
- **Weighted Recall** → Fracción de veces que no se comete error tipo I
- **Weighted F1 - Score** → Condensa info Recall y Precision

6. Resultados y Discusión

Desglose de métricas (Naive Bayes)

Intervalo	Precisión	Recall	F1-Score	Support
(0.0, 5.05)	0.80	1.00	0.89	1756
(5.05, 12.14)	0.00	0.00	0.00	289
(12.14, 21.29)	0.00	0.00	0.00	122
(21.29, 32.09)	0.42	0.57	0.49	60
(32.09, 44.25)	0.00	0.00	0.00	24
(44.25, 58.31)	0.00	0.00	0.00	16
(58.31, 75.14)	0.00	0.00	0.00	7
(75.14, 93.41)	0.00	0.00	0.00	4
(93.41, 110.25)	0.00	0.00	0.00	1

6. Resultados y Discusión



6. Resultados y Discusión

	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-Score
Naive Bayes	78.5%	62.6%	78.5%	69.69%
Random Forest	89.6%	90.9%	89.6%	90%
Decision Tree	81.2%	69.5%	81.2%	74.8%
Support Vector Machine	86.2%	81.8%	86.2%	83.3%
Multilayer Perceptron	85.3%	79.9%	85.3%	82.39%

7. Proyecciones

- Incluir diagnósticos secundarios.
- Averiguar si existe temporalidad entre los procedimientos.
- Incorporar historial clínico del paciente.





PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Predicción del tiempo de permanencia de pacientes en un hospital



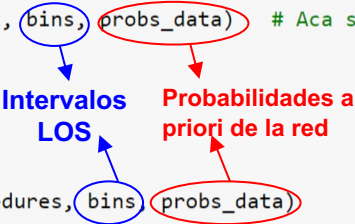
8. Anexo

Ingresar DataFrame de pacientes o solo un paciente

```
[ ] #Predicción con el dataframe X_test
    prediction = getPrediction(X_test, bins, probs_data) # Aca se testea cada entrada de X_test

#Predicción para un paciente
diagnosticos = ['I', 'U']
procedures = ['0T', '4', 'B']

predictLOSBin(diagnosticos, procedures, bins, probs_data)
```



```
↳ {(0.0, 5.053121516936446): 4.967605666756014e-05, (5.053121516936446, 12.141026503753984): 8.357049920040337e-06, (12.141026503753984, ;
    (0.0, 5.053121516936446)}
```

8. Anexo

Matriz de confusión (Naive Bayes):

- Accuracy: **78.5%**
- Weighted Precision:
62.6%
- Weighted Recall:
78.5%
- Weighted F1-score:
69.69%

1756	0	0	0	0	0	0	0	0
289	0	0	0	0	0	0	0	0
122	0	0	0	0	0	0	0	0
26	0	0	34	0	0	0	0	0
1	0	0	23	0	0	0	0	0
5	0	0	11	0	0	0	0	0
0	0	0	7	0	0	0	0	0
0	0	0	4	0	0	0	0	0
0	0	0	1	0	0	0	0	0