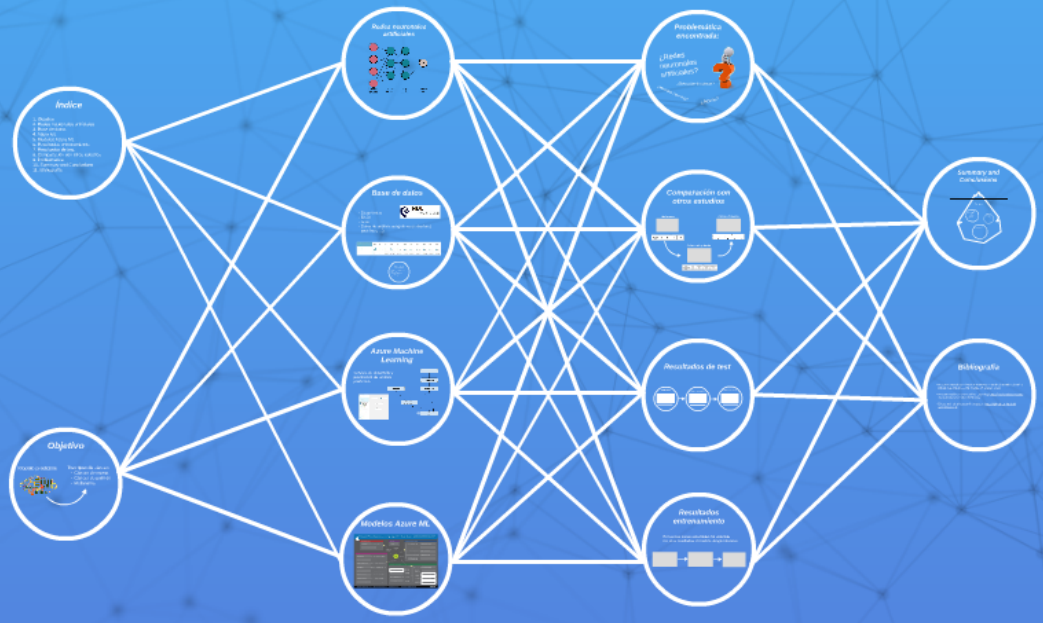
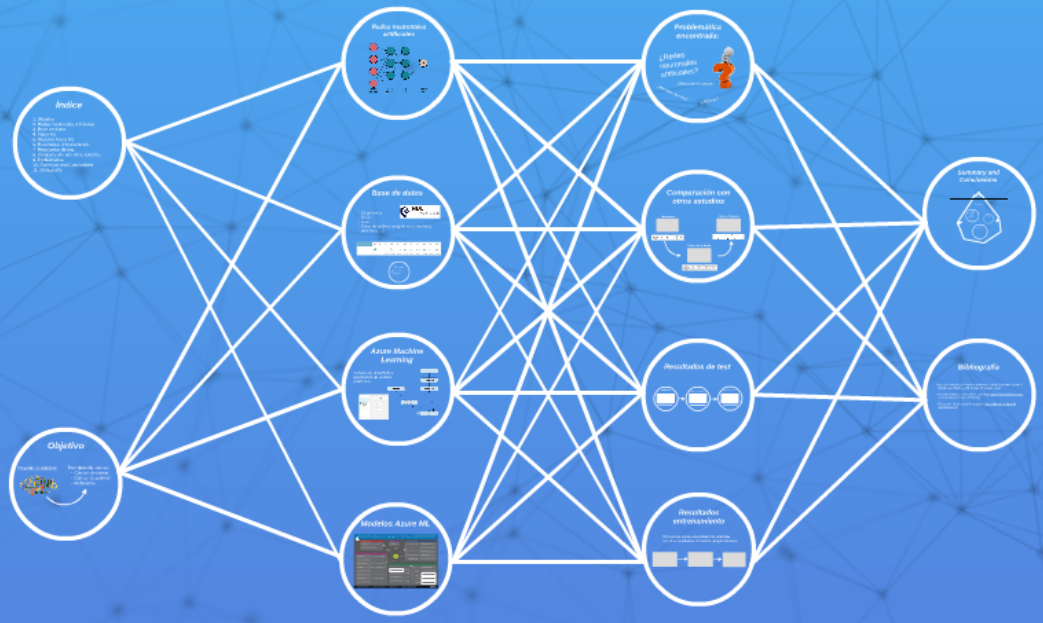
Microsoft Azure Machine Learning para detección de distintos tipos de cáncer



Autor: Óscar Trujillo Acosta
Tutor: Patricio García Báez
Co-Tutora: Carmen Paz Suárez Araujo

Microsoft Azure Machine Learning para detección de distintos tipos de cáncer

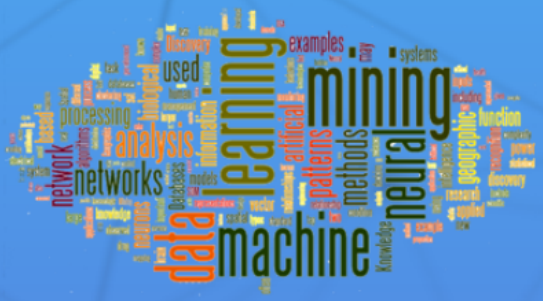


Autor: Óscar Trujillo Acosta
Tutor: Patricio García Báez
Co-Tutora: Carmen Paz Suárez Araujo

Índice

Objetivo

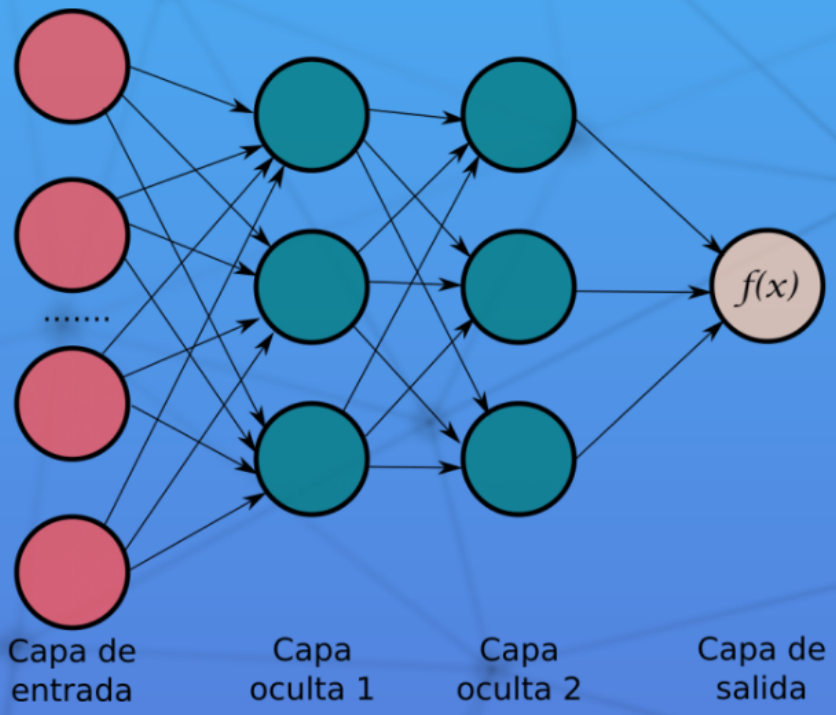
Modelo predictivo



Tres tipos de cáncer:

- Cáncer de mama
- Cáncer de pulmón
- Melanoma

Redes neuronales artificiales



Base de datos

HEAL
HEURISTIC AND EVOLUTIONARY ALGORITHMS LABORATORY

- Diagnóstico
- Edad
- Sexo
- Datos de análisis sanguíneos (colesterol, proteínas, ...)

RespiratorySystemCancer_Diagnosis	ALTER	SEX	AFP	ALT	AST	BSG1	BUN	C199	CBAA	CEA	CEOA
1	41	1	0.059995	0.425101	0.53125	0.441426	0.057692	0.240931	0.067986	0.210011	0.109463
0	67	0	0.059995	0.137652	0.151042	0.441426	0.083333	0.240931	0.06546	0.210011	0.141688

- Edad
- Sexo
- Datos de análisis sanguíneos (colesterol, proteínas, ...)

RespiratorySystemCancer_Diagnosis	ALTER	SEX	AFP	ALT	AST	BSG1	BUN	C199	CBAA	CEA	CEOA
1	41	1	0.059995	0.425101	0.53125	0.441426	0.057692	0.240931	0.067986	0.210011	0.109463
0	67	0	0.059995	0.137652	0.151042	0.441426	0.083333	0.240931	0.06546	0.210011	0.141688

Preprocesado

Eliminación de columnas y filas poco significativas

Sustitución de datos nulos

División en subconjuntos aleatorios:
Entrenamiento: 50%
Validación: 25%
Testeo: 25%

Preprocesado

Eliminación de columnas y filas poco significativas

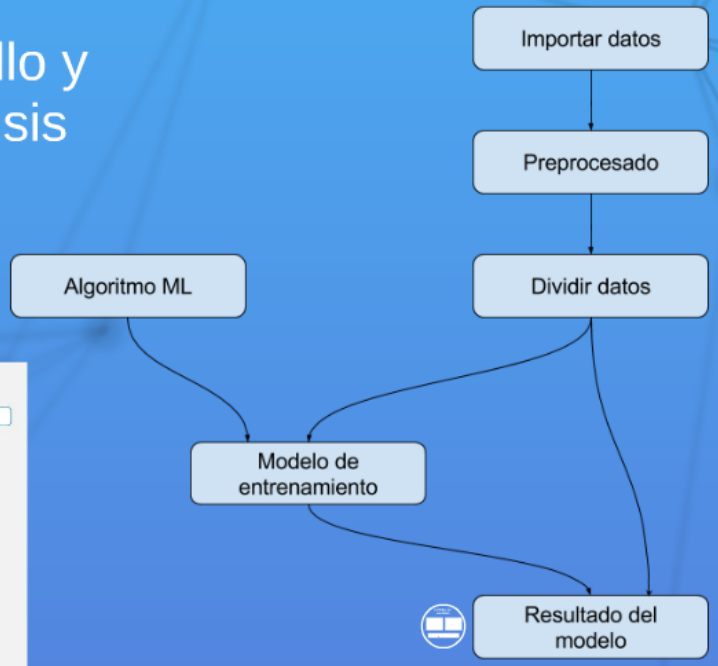
Sustitución de datos nulos

División en subconjuntos aleatorios:

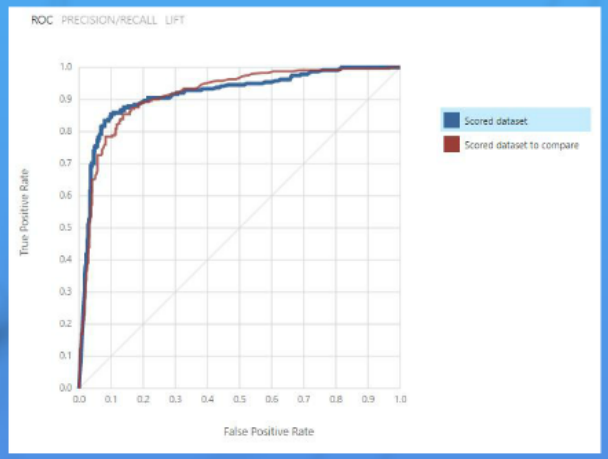
Entrenamiento:	50%
Validación:	25%
Testeo:	25%

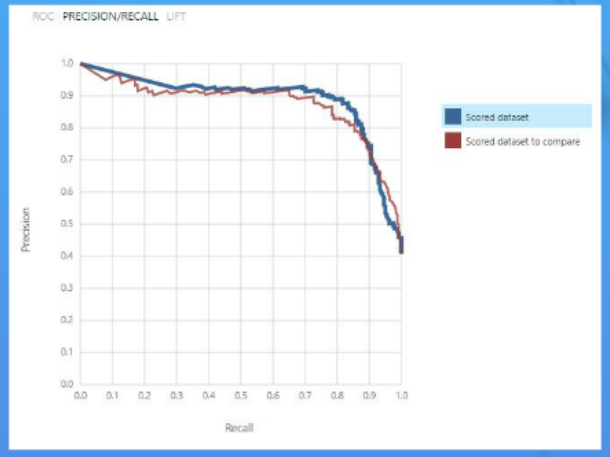
Azure Machine Learning

Servicio de desarrollo y publicación de análisis predictivo.



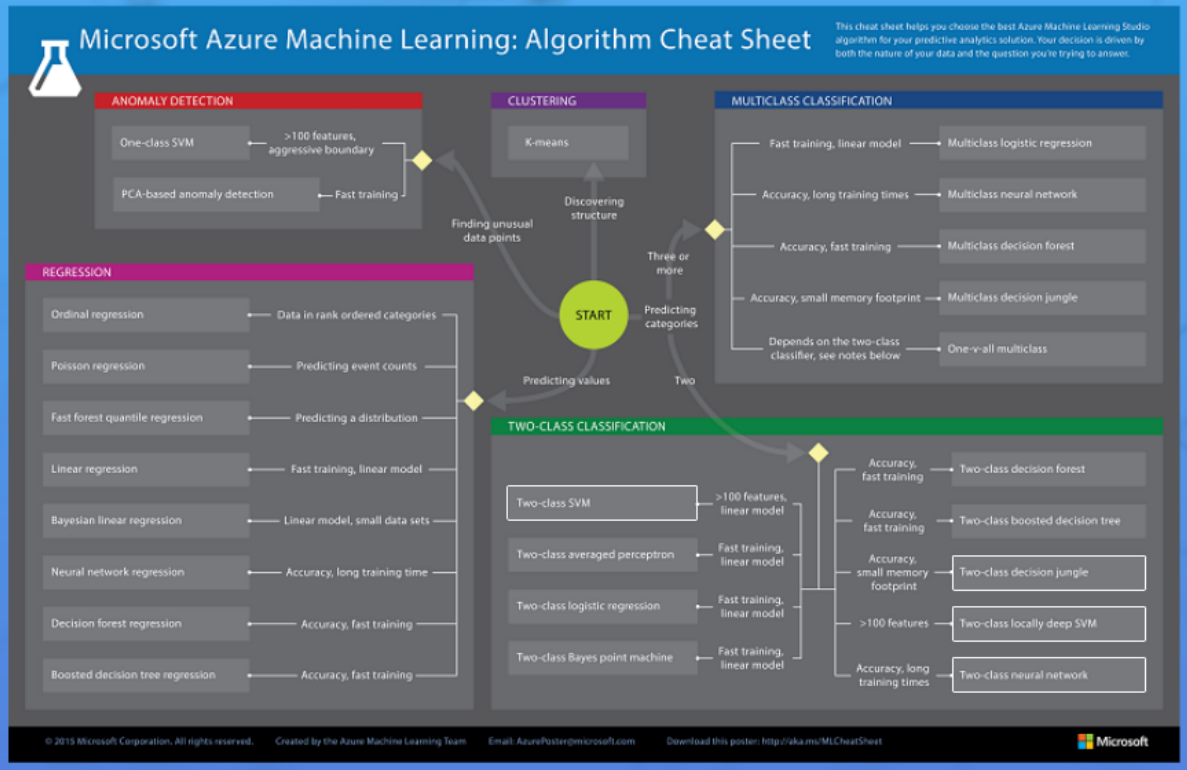
Formato del resultado



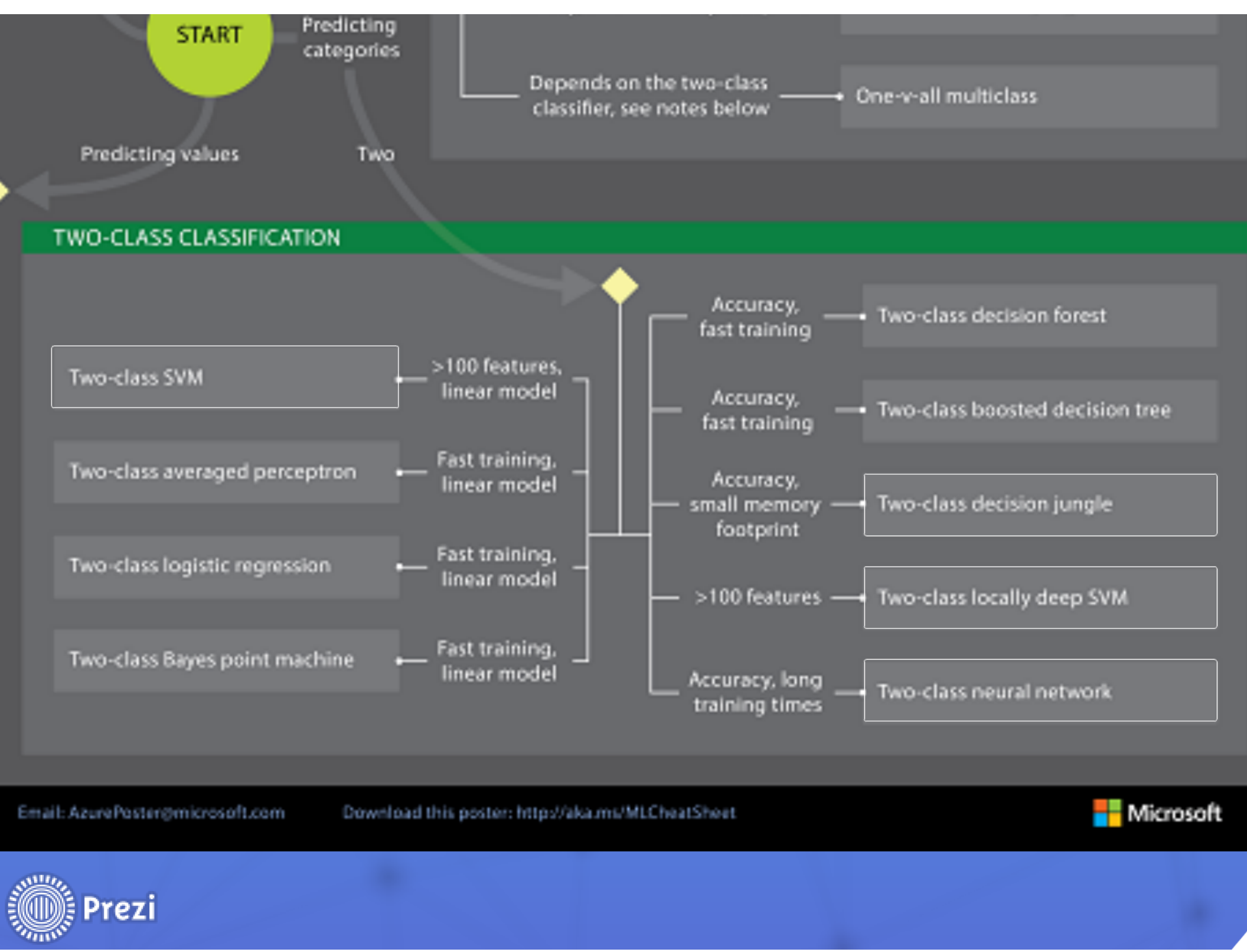


True Positive	False Negative	Accuracy	Precision	Threshold	AUC
196	45	0.883	0.891	0.5	0.916
False Positive	**True Negative**	**Recall**	**F1 Score**		
24	326	0.813	0.850		

Modelos Azure ML

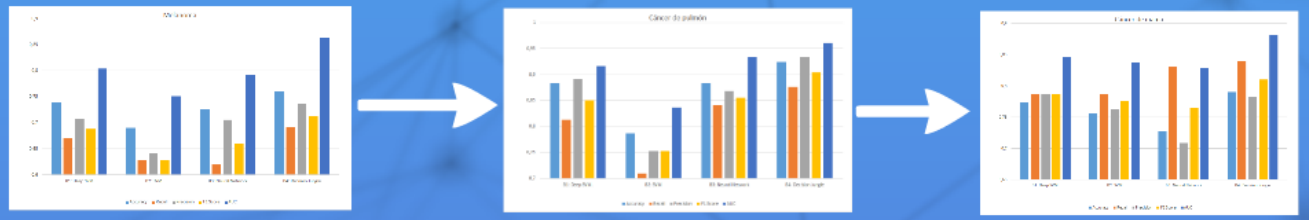


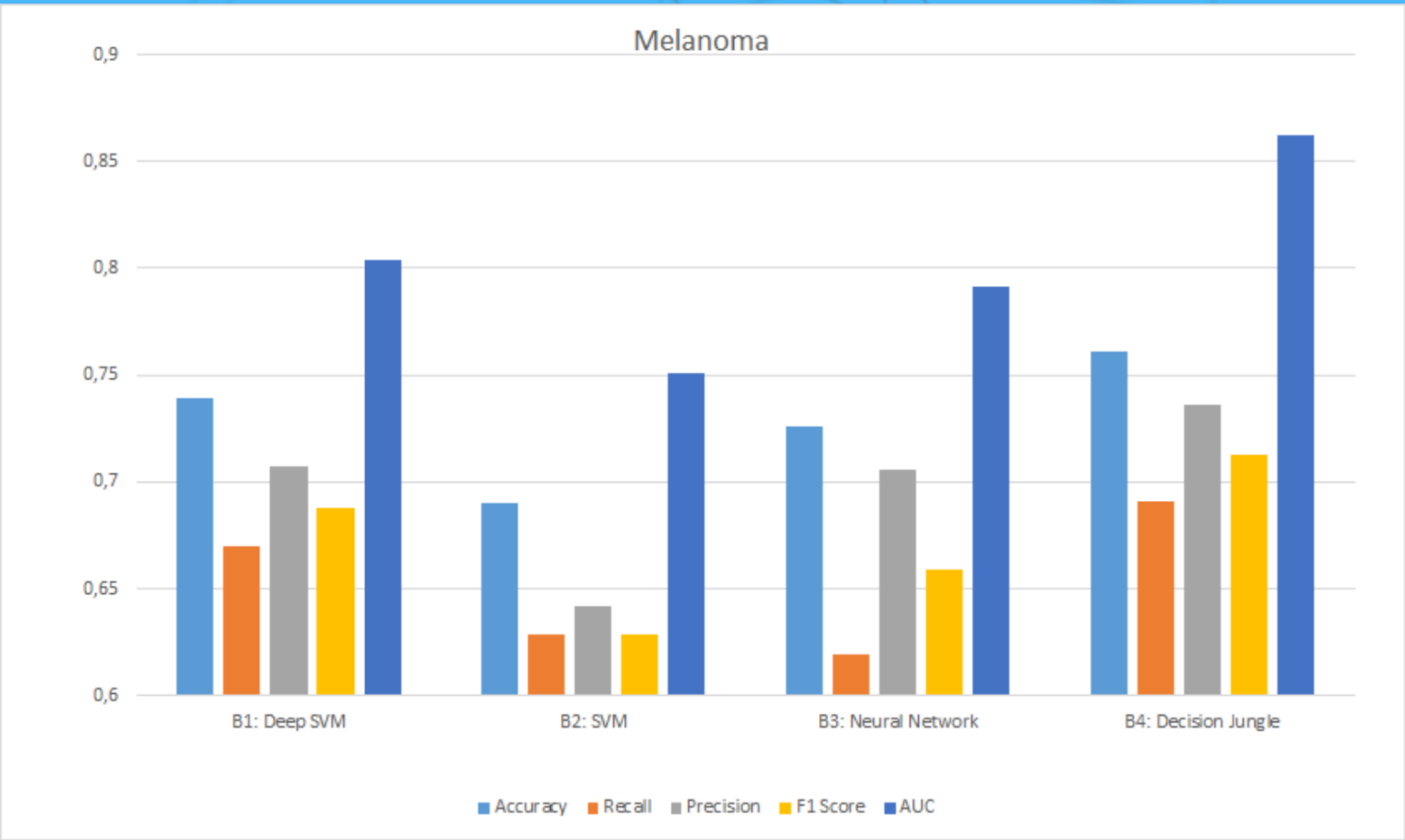
Prezi

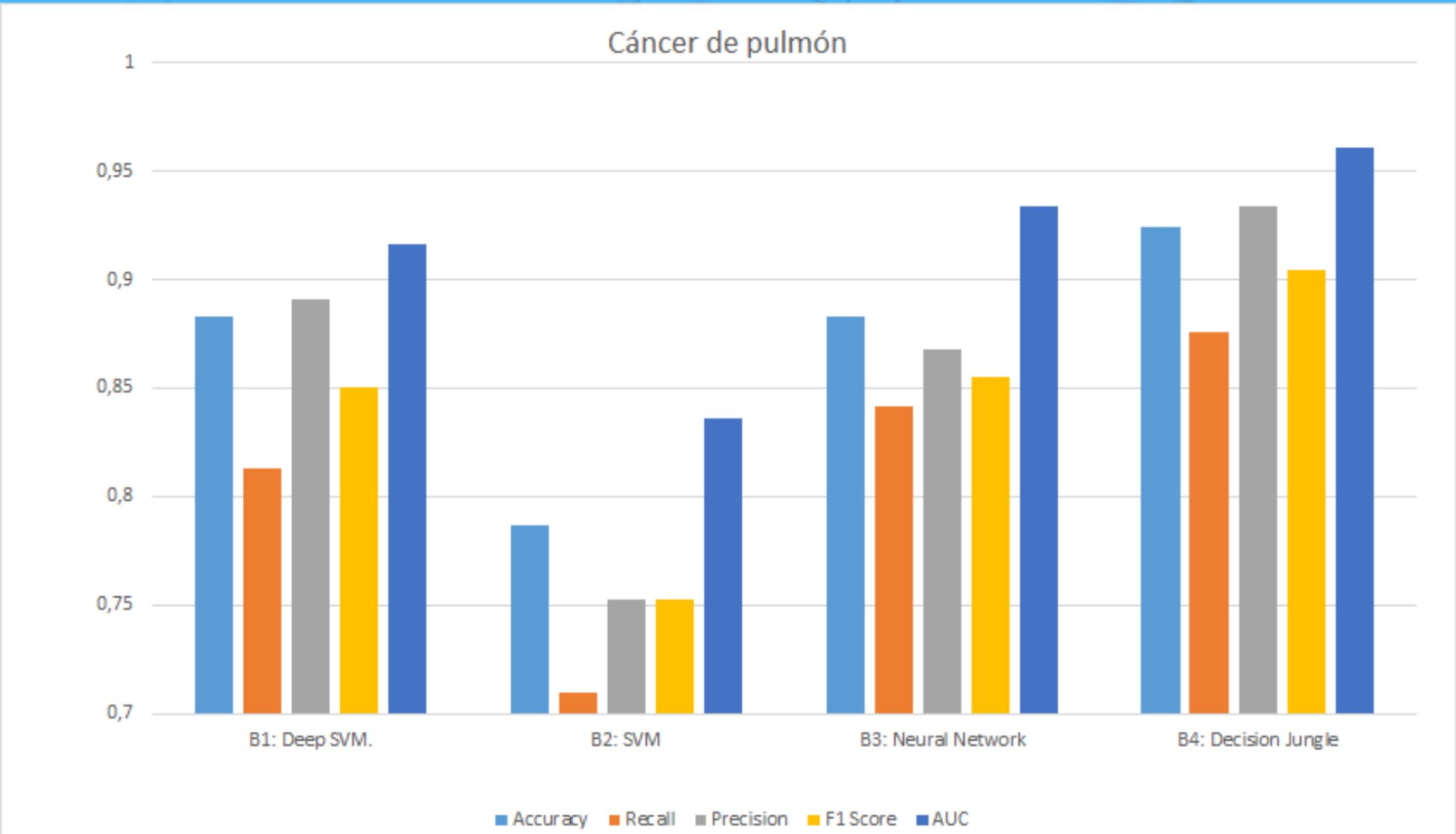

START
Predicting categories
Depends on the two-class classifier, see notes below
One-v-all multiclass
Predicting values
Two
TWO-CLASS CLASSIFICATION
Two-class SVM
>100 features, linear model
Two-class averaged perceptron
Fast training, linear model
Two-class logistic regression
Fast training, linear model
Two-class Bayes point machine
Fast training, linear model
Accuracy, fast training
Two-class decision forest
Accuracy, fast training
Two-class boosted decision tree
Accuracy, small memory footprint
Two-class decision jungle
>100 features
Two-class locally deep SVM
Accuracy, long training times
Two-class neural network
Email: AzurePoster@microsoft.com
Download this poster: http://aka.ms/MLCheatSheet
Microsoft

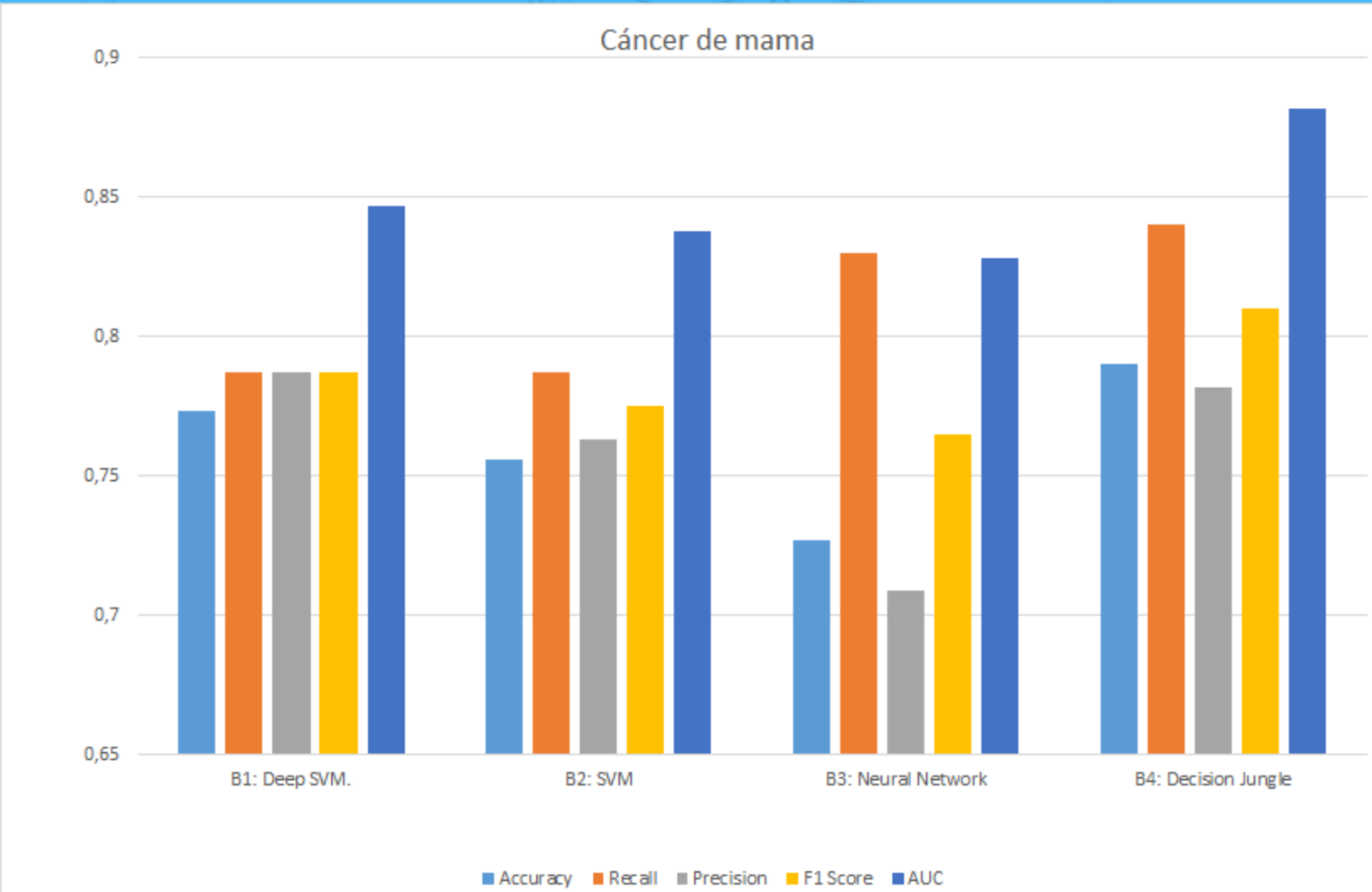
Resultados entrenamiento

En los tres casos estudiados ha obtenido mejores resultados el modelo Jungle Decision



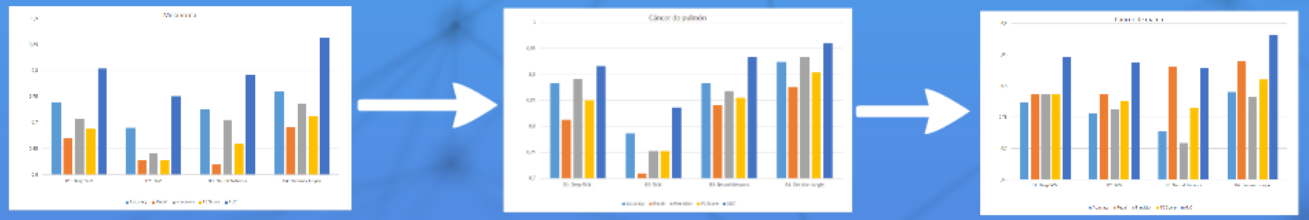

Melanoma
0,9
0,85
0,8
0,75
0,7
0,65
0,6
B1: Deep SVM
B2: SVM
B3: Neural Network
B4: Decision Jungle
Accuracy
Recall
Precision
F1 Score
AUC


Cáncer de pulmón
1
0,95
0,9
0,85
0,8
0,75
0,7
B1: Deep SVM.
B2: SVM
B3: Neural Network
B4: Decision Jungle
Accuracy
Recall
Precision
F1 Score
AUC


Cáncer de mama
0,9
0,85
0,8
0,75
0,7
0,65
B1: Deep SVM.
B2: SVM
B3: Neural Network
B4: Decision Jungle
Accuracy
Recall
Precision
F1 Score
AUC

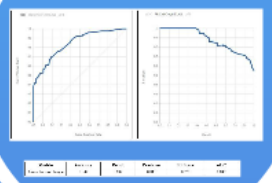
Resultados entrenamiento

En los tres casos estudiados ha obtenido mejores resultados el modelo Jungle Decision

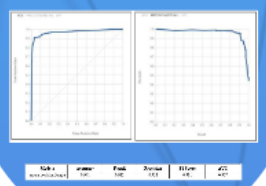


Resultados de test

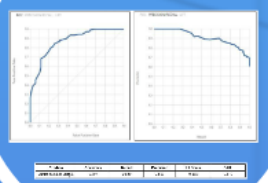
Melanoma



Cáncer de pulmón

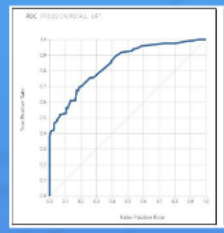


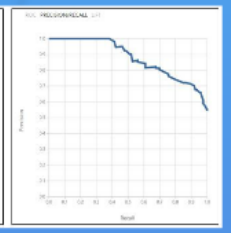
Cáncer de mama



Resultados de test

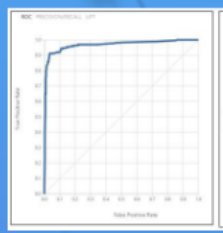
Melanoma

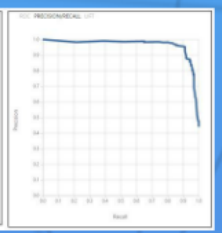




Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Árbol decisión Jungla	0.748	0.698	0.811	0.751	0.842

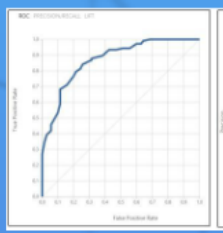
Cáncer de pulmón

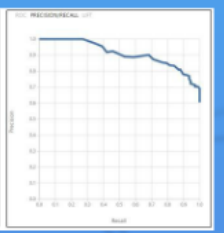




Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Árbol decisión Jungla	0.811	0.813	0.870	0.831	0.907

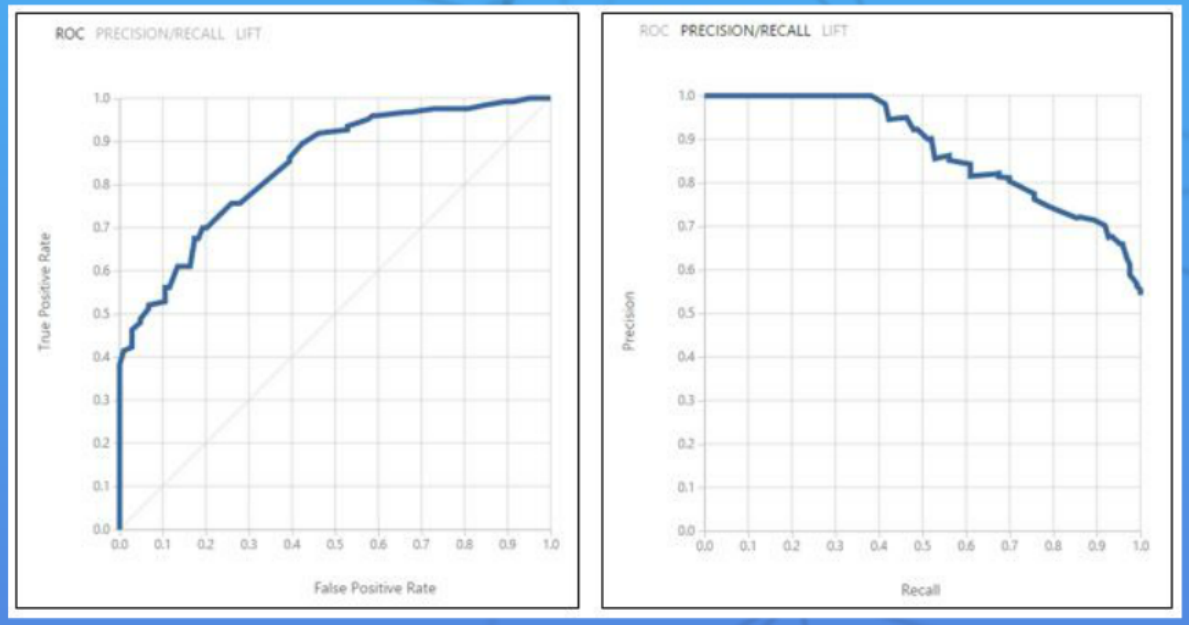
Cáncer de mama





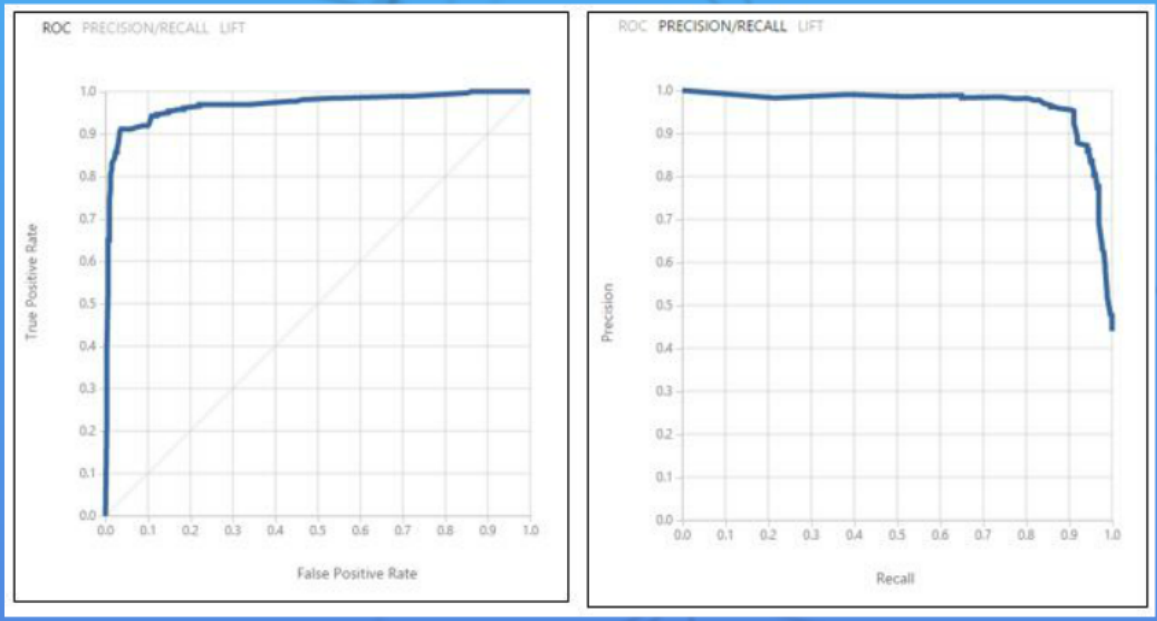
Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Árbol decisión Jungla	0.797	0.869	0.809	0.838	0.873

Melanoma



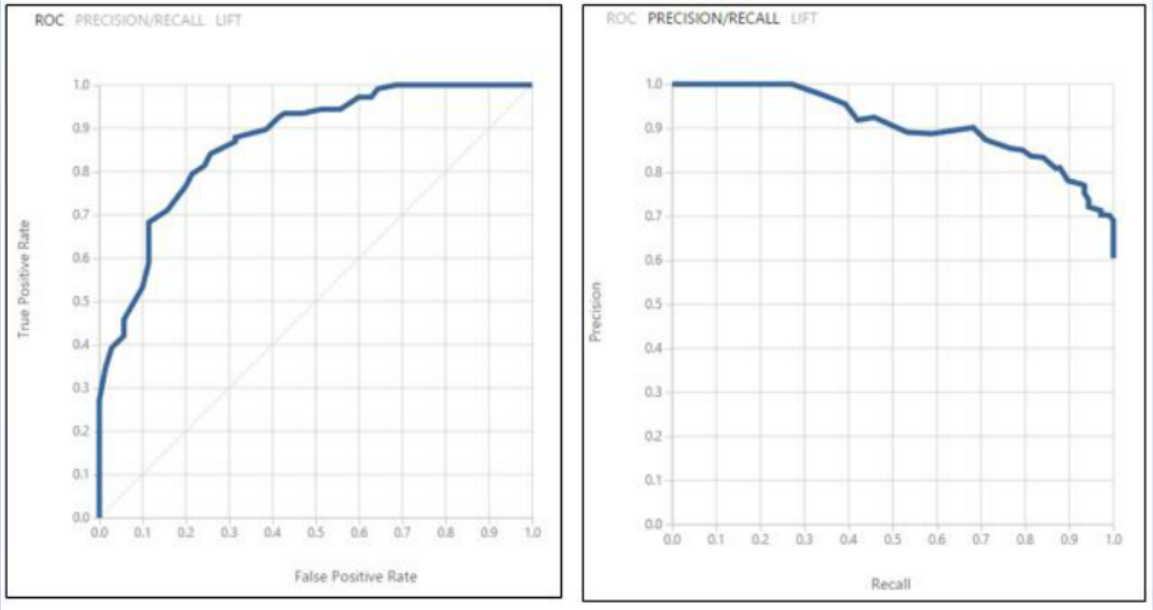
Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Azure decision Jungle	0.749	0.699	0.811	0.751	0.842

Cáncer de pulmón



Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Azure decision Jungle	0.941	0.912	0.952	0.931	0.967

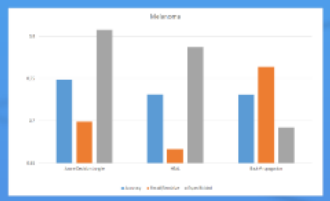
Cáncer de mama



Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Azure decision Jungle	0.797	0.869	0.809	0.838	0.870

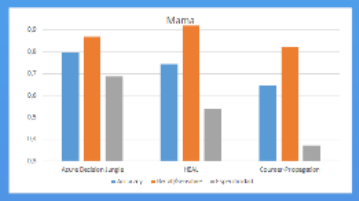
Comparación con otros estudios

Melanoma



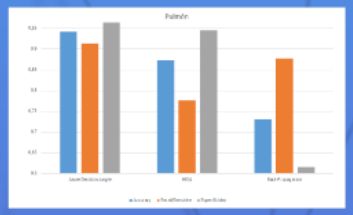
Resultados	Testeo			
	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas
Azure Decision Jungle	0.740	0.693	0.8076	0.1171
HEAL	0.7315	0.6667	0.7876	0
Back-Propagation	0.7510	0.7640	0.6920	0.1313

Cáncer de mama



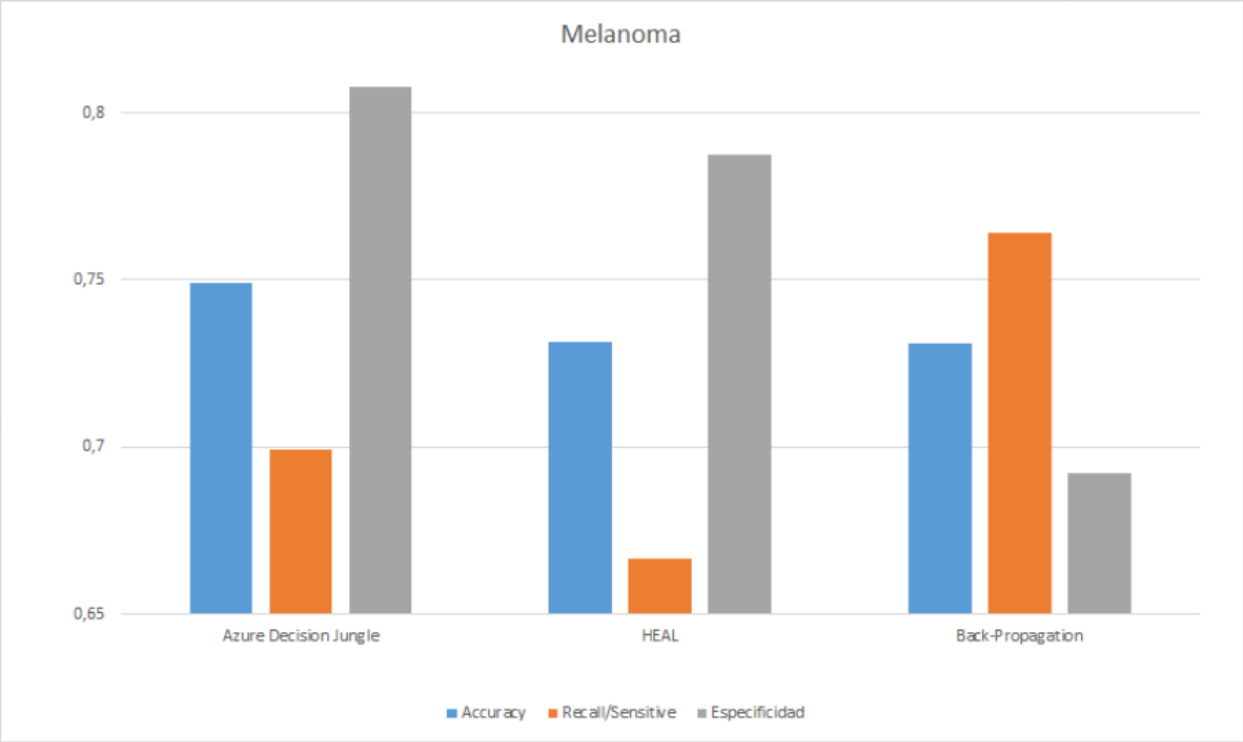
Resultados	Testeo			
	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas (%)
Azure Decision Jungle	0.787	0.869	0.6857	0.0127
HEAL	0.7456	0.9188	0.5376	0
Counter Propagation	0.644	0.822	0.371	0.0127

Cáncer de pulmón



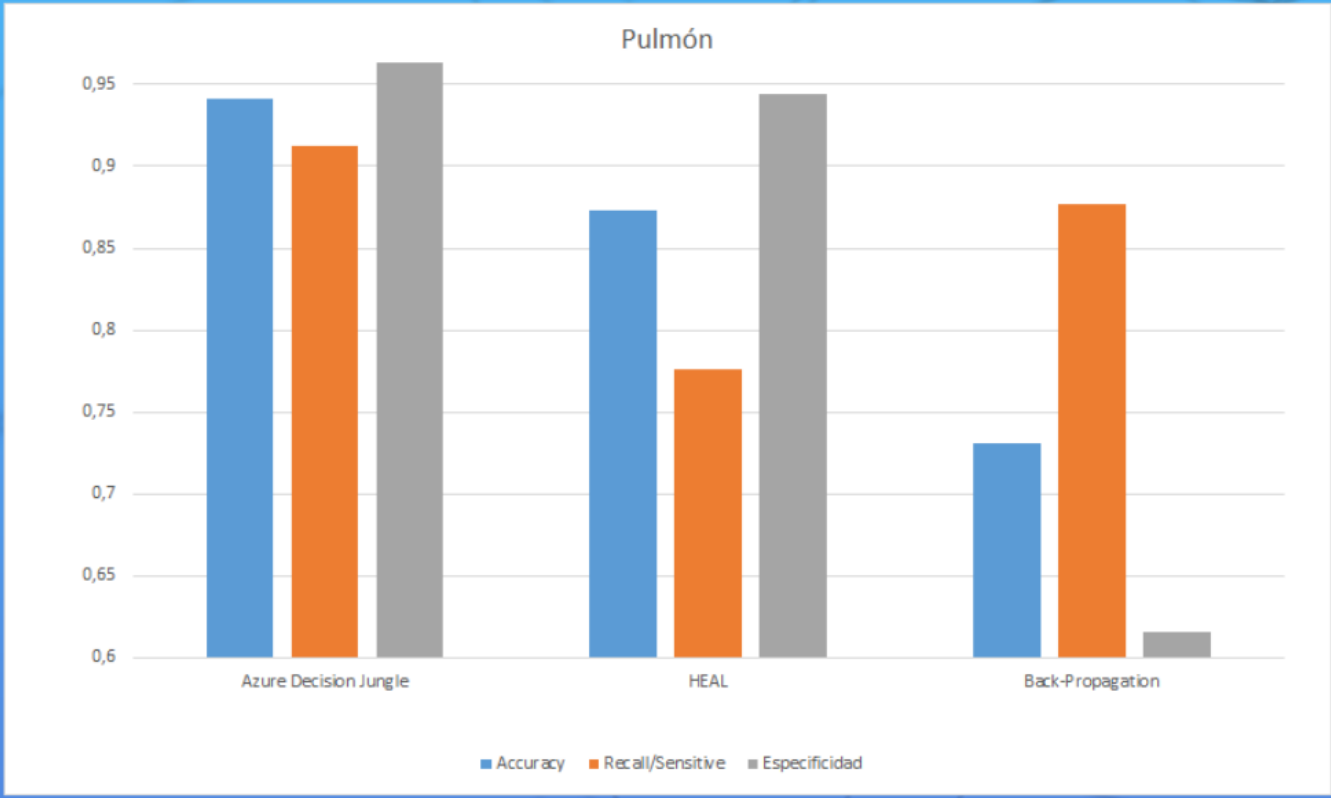
Resultados	Testeo			
	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas (%)
Azure Decision Jungle	0.941	0.912	0.9637	0.1557
HEAL	0.8750	0.7761	0.9437	0
Back-Propagation	0.7210	0.8170	0.6160	0.1557

Melanoma



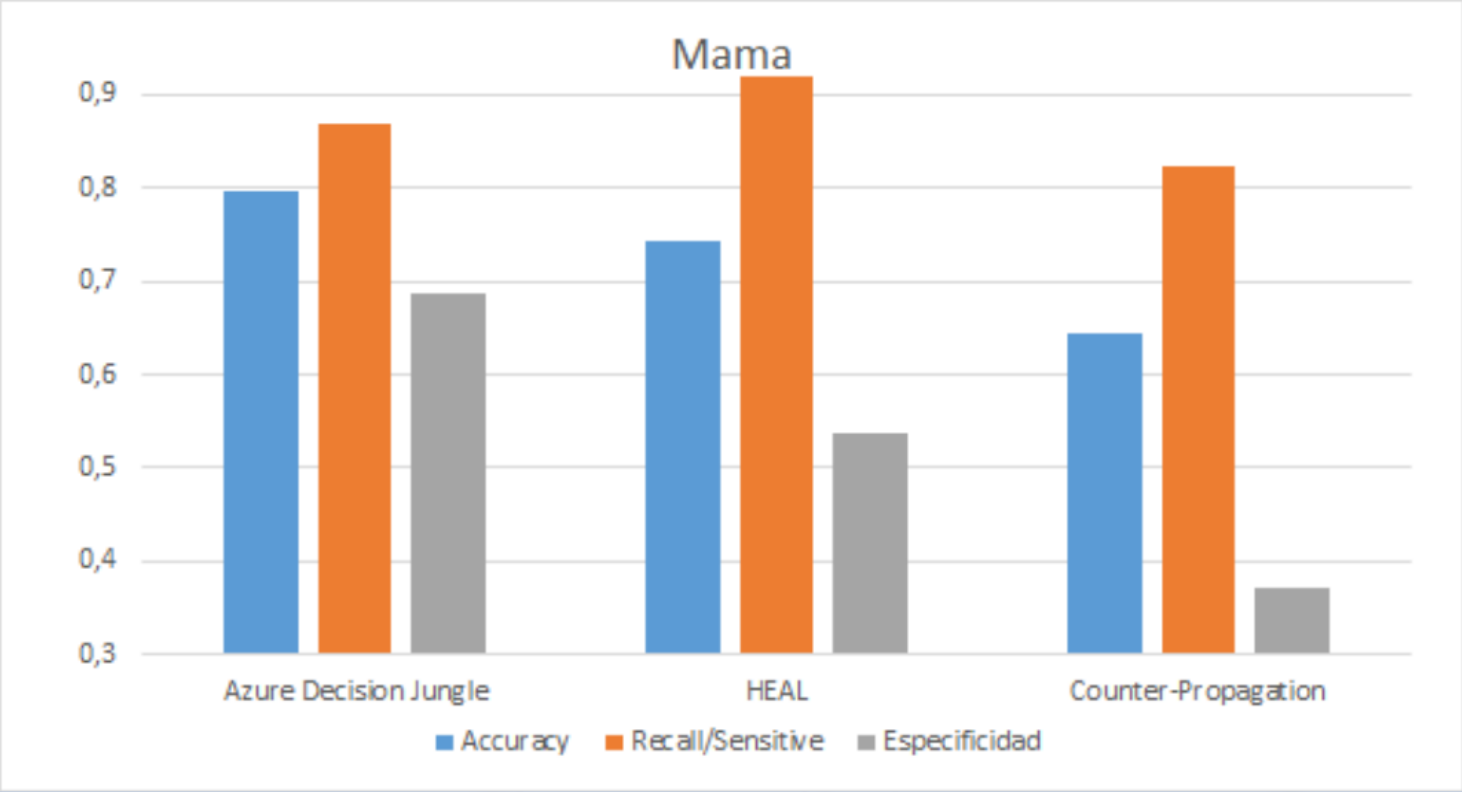
	Testeo			
Resultados	**Accuracy**	**Recall/Sensitive**	**Especificidad**	**Filas eliminadas**
Azure Decision Jungle	**0.749**	0.699	**0.8076**	0.1171
HEAL	0.7315	0.6667	0.7876	0
Back-Propagation	0.7310	**0.7640**	0.6920	0. 1171

Cáncer de pulmón



	Testeo			
Resultados	**Accuracy**	**Recall/Sensitive**	**Especificidad**	**Filas eliminadas (%)**
Azure Decision Jungle	**0.941**	**0.912**	**0.9637**	0.1557
HEAL	0.8730	0.7761	0.9437	0
Back-Propagation	0.7310	0.8770	0.6160	0. 1557

Cáncer de mama



	Testeo			
Resultados	**Accuracy**	**Recall/Sensitive**	**Especificidad**	**Filas eliminadas (%)**
Azure Decision Jungle	**0.797**	0.869	**0.6857**	0.0127
HEAL	0.7436	**0.9188**	0.5370	0
Counter-Propagation	0.644	0.822	0.371	0.0127

Problemática encontrada:

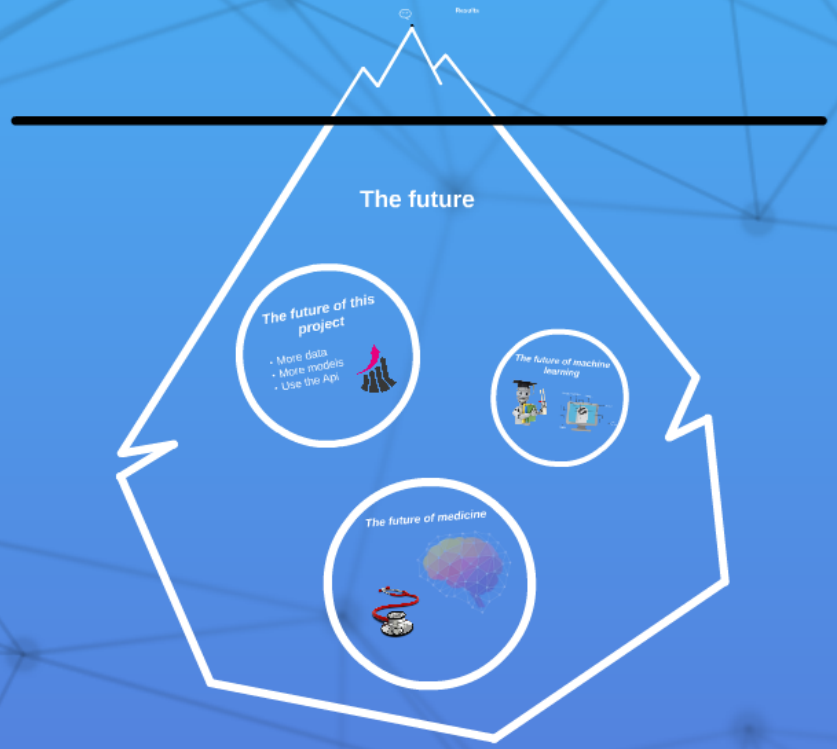
¿Redes neuronales artificiales?

¿Detección de cáncer?

¿Machine Learning?

¿Azure?

Summary and Conclusions



Results

The future

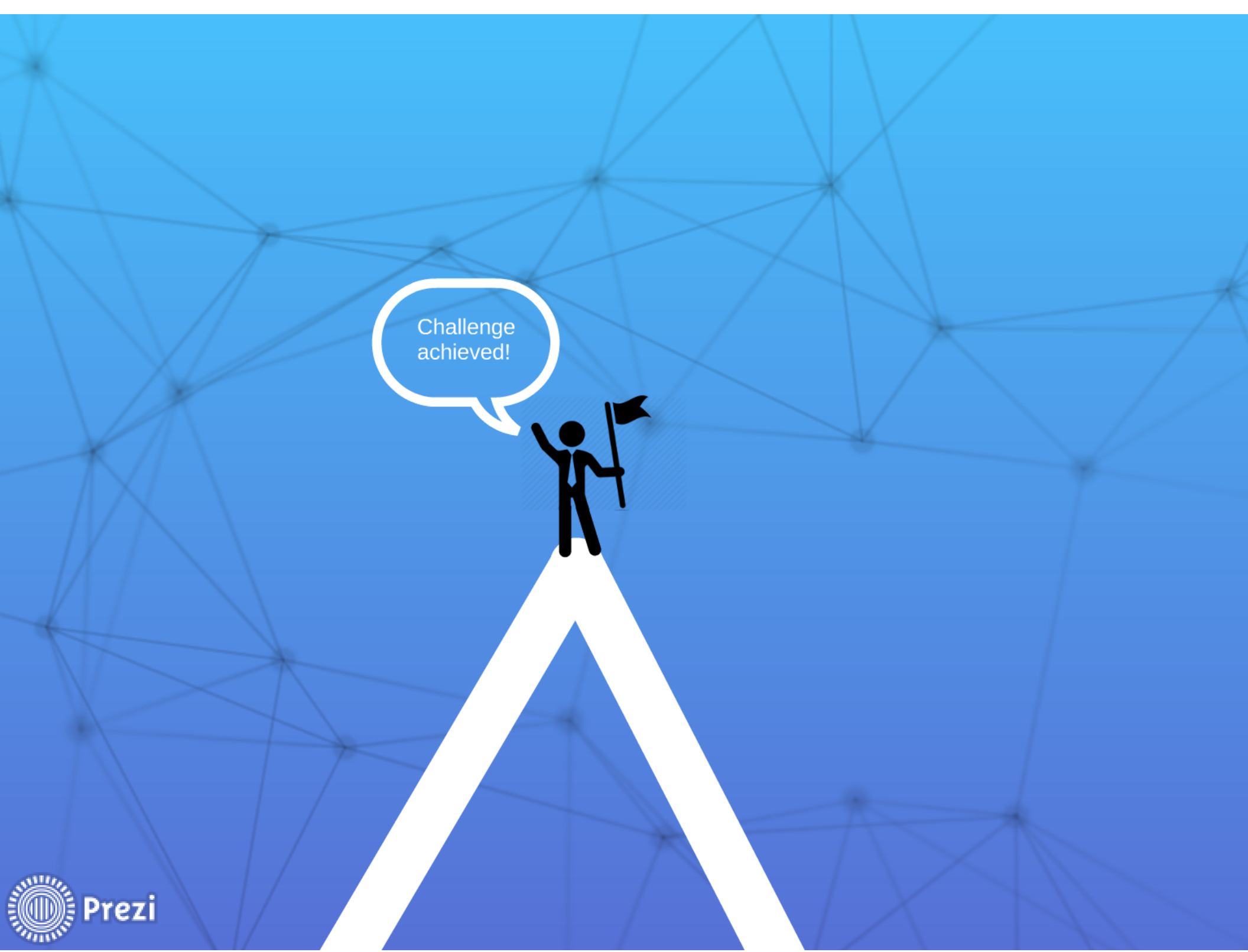
The future of this project

- More data
- More models
- Use the Api

The future of machine learning

The future of medicine

Results


Challenge achieved!

Results

The future

The future of this project

- More data
- More models
- Use the Api

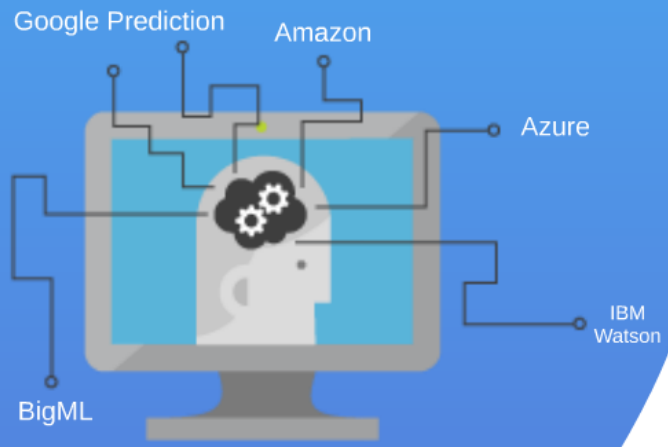
The future of machine learning

The future of medicine

The future of this project

- More data
- More models
- Use the Api

The future of machine learning

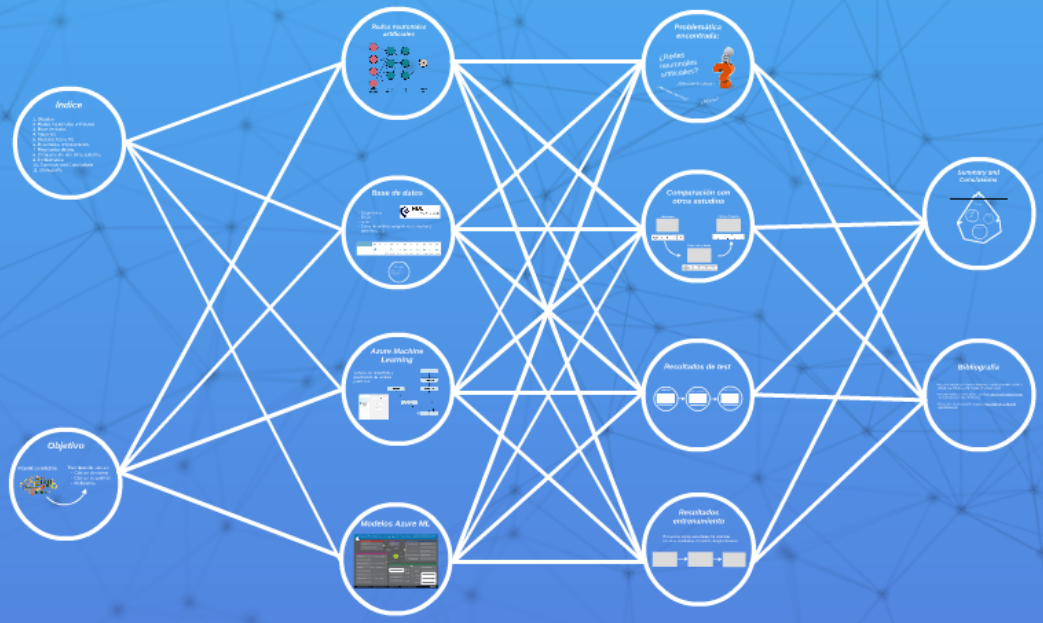


The future of medicine

Bibliografía

- Redes Neuronales y Sistemas Borrosos – Bonifacio Martín del Brío y Alfredo Sanz Molina – Ed. Ra-Ma, 2ª edición, 2001
- Documentación Azure Machine Learning. https://msdn.microsoft.com/en-us/library/azure/dn578280,aspx
- Enlace GitHub al trabajo fin de grado. https://github.com/etsiiull/azuremedicalnn

Microsoft Azure Machine Learning para detección de distintos tipos de cáncer



Autor: Óscar Trujillo Acosta
Tutor: Patricio García Báez
Co-Tutora: Carmen Paz Suárez Araujo