



PREDICTION DE PERFORMANCE DES ETUDIANTS AVEC LES METHODES D'AUTO APPRENTISSAGE

(MAKİNE ÖĞRENMESİ METODLARIYLA ÖĞRENCİ PERFORMANS TAHMİNİ)

Çağrı Münyas, Dr. Sultan Turhan

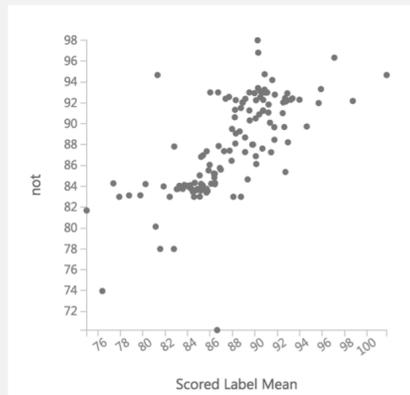
Département de Génie Informatique
Faculté d'Ingénierie et de Technologie, Université Galatasaray

Le but du Projet

- Collecter les données collectées par différents agents des institutions académiques
- Utiliser l'apprentissage automatique pour créer une application prédictive afin d'améliorer les performances des étudiants en classe, en 2 étapes :
 - Créer un modèle de classification pour prédire le taux de réussite des étudiants.
 - Regrouper les étudiants similaires et assigner les nouvelles étudiantes aux groupes
 - Coupler les étudiants avec les notes plus basses et plus hautes dans le même groupe

Comparaison des modèles de Régression et de Classification

- Lorsque la valeur cible est numérique :



Les Résultats du modèle de Régression Linéaire

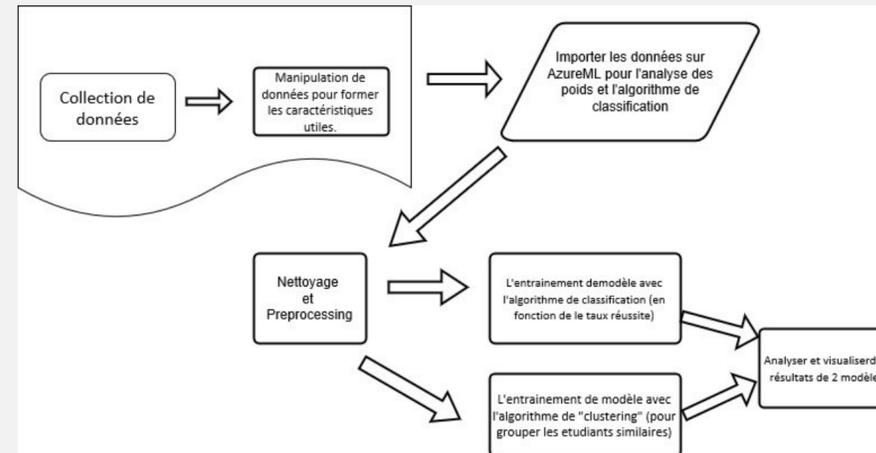
- lorsque la valeur cible est catégorique:

		Predicted Class		
		1	2	3
Actual Class	1	81.1%	16.2%	2.7%
	2	20.8%	68.8%	10.4%
	3	5.0%	12.5%	82.5%

Les Résultats du modèle de Classification

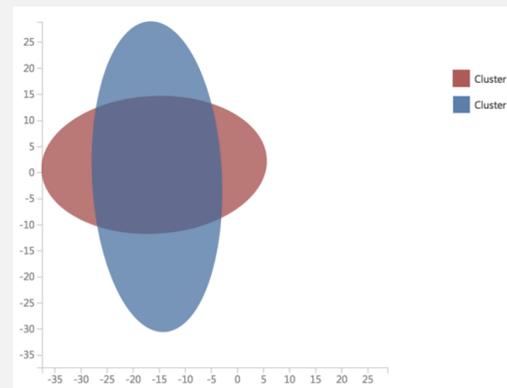
Les Etapes du Projet

- Le diagramme des étapes d'application du projet :



Les Problèmes rencontrés et les bloqueurs

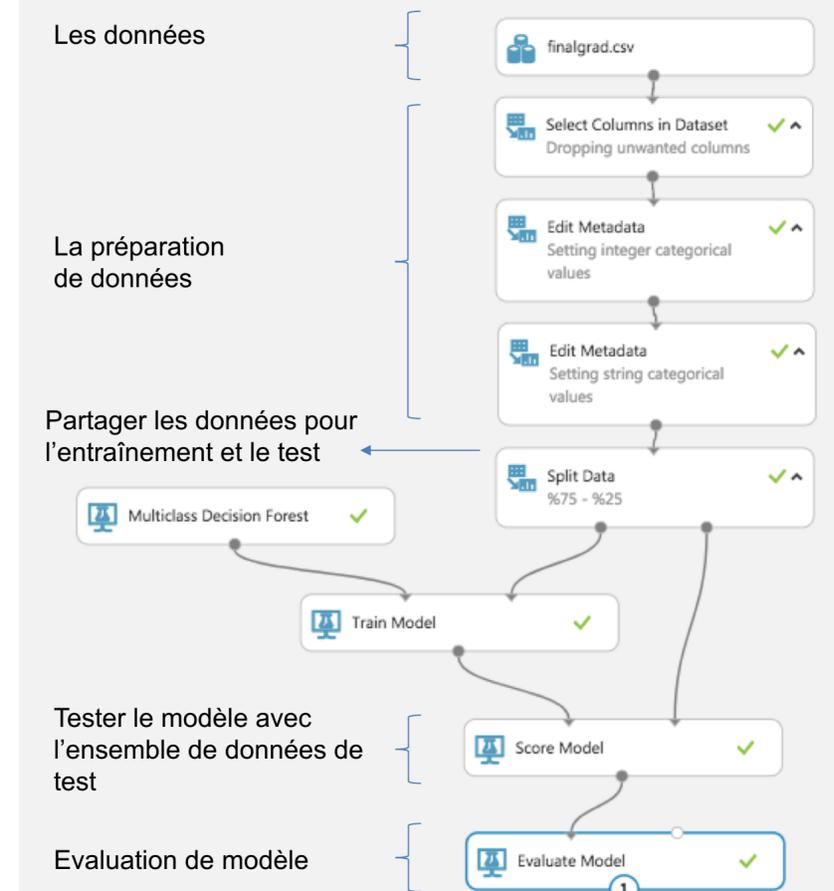
- Les groupes après l'affectation des données avec l'entraînement de clustering:



- Ce modèle a échoué. Comme on peut le voir dans la Figure, les groupes se chevauchent. Ceci est causé par la quantité insuffisante de données.
- Le plus grand défi était la collecte de données parce que personne n'est prêt à donner des données à cause de la nouvelle loi sur la protection des données personnelles (KVKK). Le problème est résolu en demandant des données personnelles qui devraient être fournies par les instituts Académiques, directement aux étudiants dans l'enquête
- Le deuxième problème était les données déséquilibrées. La donnée biaisée provoque des résultats inattendus. En utilisant la méthode SMOTE (Synthetic Minority Oversampling Technique), on équilibre l'ensemble de données avec de nouvelles exemples synthétiques.

Les Etapes de Préparation et Entraînement sur AzureML

- Les étapes de préparation de données et l'entraînement sur AzureML Studio



Conclusion

Dans cette simulation :

- On a créé les modèles de Régression et de Classification avec le taux de réussite acceptable, pour prédire la réussite des étudiants.
- On a échoué de regrouper ces étudiants dans un sens utilisable, à cause de la quantité insuffisante de données.

Au Futur:

- Si on peut recueillir plus de données, on peut regrouper les étudiants et obtenir de meilleurs résultats des modèles de Régression, comme on aura plus d'échantillon pour l'entraînement.
- L'application de prédire les notes des étudiants et les coupler avec les étudiants similaires qui peuvent obtenir une meilleure note selon les prédictions pour l'amélioration de succès global