République Algérienne Démocratique et Populaire Ministère de l'Enseignement Supérieur et de la Recherche Scientifique Université d'Alger 1



Faculté des sciences - Département Informatique

Fillière : Informatique - Spécialité : SI

Prédiction Des Revenus Mensuels En Utilisant L'apprentissage Automatique Non Supervisé (Le Clustering).



Réalisé par : Nouasri Lina Manel Larkem Kahina Bahia

Encadré par : M. BENTALEB Abdelfettah

Mme. TATACHAK Fatma

Zohra Amina

Remerciements

Tout d'abord, nous souhaitons remercier chaleureusement mes encadrants de stage chez Djezzy, BENTALEB Abdelfettah et TATACHAK Fatma Zohra Amina, pour leur accueil, leur confiance et leur encadrement tout au long de cette expérience enrichissante. Leurs conseils avisés et leur expertise ont été déterminants dans la réalisation de ce projet

Nous souhaitons également remercier l'entreprise Djezzy pour nous avoir accueillis et offert cette opportunité enrichissante. Nous tenons à souligner la qualité de la formation reçue à Djezzy, qui nous a fourni les bases théoriques et pratiques essentielles à la réalisation de ce projet.

Ce stage a été une expérience formatrice exceptionnelle, tant sur le plan professionnel que personnel, et nous sommes reconnaissantes envers tous ceux qui y ont contribué.

Résumé

Au cours de notre stage chez Djezzy, nous avons développé un modèle prédictif utilisant le clustering pour estimer les revenus mensuels de l'entreprise. Le clustering, une technique d'apprentissage non supervisé, joue un rôle crucial dans la segmentation des données en groupes homogènes. Cette segmentation permet de mieux comprendre les tendances et les comportements au sein des données, offrant une vue plus claire et précise des divers segments du marché.

L'application du clustering est particulièrement utile dans un contexte commercial où les prévisions précises sont essentielles pour ajuster les stratégies en fonction des variations des conditions et des facteurs externes. En segmentant les données selon des caractéristiques similaires, nous avons pu révéler des patterns récurrents et identifier des segments de marché spécifiques. Cela a permis de fournir des insights précieux pour la planification stratégique.

Grâce à cette technique, nous avons pu offrir des prévisions des revenus qui aideront Djezzy à adapter ses décisions et optimiser ses opérations. Ainsi, le clustering a démontré son importance en améliorant la prise de décision basée sur les données et en fournissant des outils pour une gestion plus efficace et réactive aux évolutions du marché.données.

Table des matières

Ta	able des matières	6
Ta	able des figures	7
Li	ste des acronymes	8
1	Présentation de l'organisme d'accueil 1 Introduction à Djezzy	9 9
2	Contexte du stage	10
3	Descriptions des missions Nettoyage des données et visualisation pour l'étude du dataset : Entraînement du modèle sans revenus pour prédire les revenus : Prédiction des revenus d'un nouveau mois sans données de revenus et calcul de l'erreur :	12 12 13 15
4	Conclusion	16

Table des figures

3.1	Analyse de la répartition des revenus par jour et par événement (La couleur						
	des points est basée sur Day_coded et la taille des points est basée sur le jour						
	du mois)	13					
3.2	Visualisation des clusters et de leurs centroids	14					
3.3	Les différentes distributions des revenus pour chaque cluster	14					

Liste des acronymes

CSV Comma Separated Values

EDA Exploratory Data Analysis

MAE Mean Absolut Error



Présentation de l'organisme d'accueil

1 Introduction à Djezzy

Djezzy, fondée en 2001, est l'un des principaux opérateurs de télécommunications en Algérie. La société offre une gamme complète de services de téléphonie mobile, d'internet et de solutions numériques, avec une mission de connecter les Algériens à un monde numérique en constante évolution. En tant que pionnier dans l'innovation technologique, Djezzy s'engage à fournir des services de qualité à ses clients tout en contribuant au développement économique et social du pays.

1.1 Le département Big Data

Dans le cadre de notre stage, nous avons eu l'opportunité de découvrir le département Big Data, un pilier stratégique au sein de l'entreprise. Ce département joue un rôle essentiel dans la collecte, le traitement et l'analyse de grandes quantités de données générées par les activités de l'entreprise. Il utilise des technologies avancées pour transformer les données en informations exploitables, aidant ainsi à la prise de décision stratégique, à l'optimisation des services, et à la personnalisation des offres pour les clients. Grâce à des outils analytiques et à des équipes de data scientists, le département contribue à l'innovation et à l'amélioration continue des performances de l'entreprise.

CHAPITRE

Contexte du stage

Dans le cadre de notre stage au sein de Djezzy, nous avons été intégrés à l'équipe Big Data, où notre mission principale consistait à effectuer du clustering pour prédire les revenus mensuels. Cette approche de segmentation a pour but de mieux comprendre les comportements et les tendances qui influencent les revenus, en tenant compte de diverses conditions telles que les jours fériés, les cycles de rechargement, les événements spéciaux, et les actions commerciales menées par l'entreprise.

L'objectif principal de cette analyse est de fournir à Djezzy des prévisions précises des revenus pour chaque mois, permettant ainsi aux équipes décisionnelles d'adapter leurs stratégies commerciales et marketing en conséquence. Par exemple, en identifiant les périodes de forte et de faible performance, l'entreprise peut ajuster ses promotions, optimiser ses campagnes publicitaires, et planifier des actions ciblées pour maximiser les revenus.

Pour atteindre cet objectif, nous avons utilisé des techniques de clustering, telles que le KMeans, pour regrouper les données en segments distincts. Chaque cluster représente un profil de revenus spécifique associé à certaines conditions, comme les jours de la semaine, les événements saisonniers, et les types d'actions commerciales. Ces regroupements permettent à Djezzy de cibler ses efforts de manière plus efficace, en élaborant des stratégies adaptées à chaque profil identifié.

Les résultats obtenus serviront donc comme une base pour la mise en œuvre de nouvelles stratégies commerciales, en optimisant les ressources et en répondant de manière plus agile aux fluctuations du marché. Notre travail de prédiction des revenus

vise à transformer les données en un levier stratégique, soutenant ainsi les ambitions de croissance et d'innovation de Djezzy.



Descriptions des missions

Notre mission principale était de développer un modèle prédictif basé sur le clustering pour estimer les revenus mensuels en tenant compte de différentes conditions et facteurs externes. Pour atteindre cet objectif, nous avons suivi un processus structuré en plusieurs étapes détaillées.

1 Nettoyage des données et visualisation pour l'étude du dataset :

La première étape a été de préparer le dataset pour l'analyse. Notre base de données comprenait diverses colonnes telles que Date, Day, Revenu, Event_coded, Action_Businnes, Day_of_month, Week, Cycle_de_rechargement, Month, et Jour_ferie. Nous avons commencé par le nettoyage des données en traitant les valeurs manquantes, en vérifiant les incohérences, et en normalisant les variables pour assurer leur utilisabilité dans le modèle.

Pour la préparation des données, nous avons utilisé One Hot Encoder pour transformer les variables catégorielles en variables binaires, facilitant ainsi leur utilisation par le modèle. Les données ont ensuite été normalisées à l'aide de Standard Scaler, qui a permis de centrer et de réduire les valeurs des features afin d'améliorer la performance du clustering.

Ensuite, nous avons réalisé une analyse exploratoire des données (EDA) à l'aide de visualisations comme des scatter plots, des plots et des histogrammes. Par exemple,

nous avons observé l'influence des jours fériés (Jour_ferie) et des cycles de rechargement (Cycle_de_rechargement) sur les revenus. Cette phase nous a aidés à comprendre les schémas sous-jacents et à identifier les variables clés pour la prédiction.

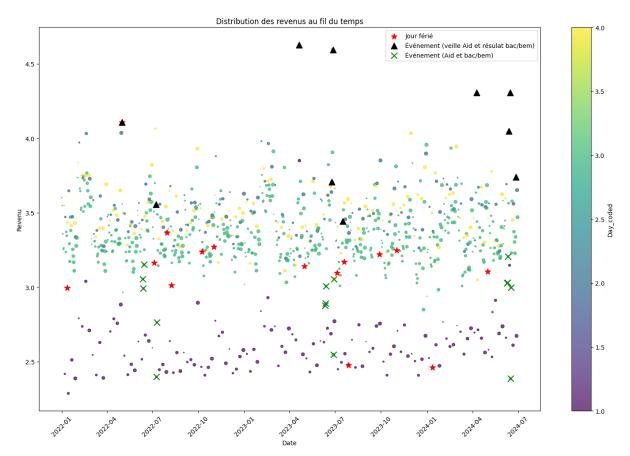


Fig. 3.1 Analyse de la répartition des revenus par jour et par événement (La couleur des points est basée sur Day_coded et la taille des points est basée sur le jour du mois)

2 Entraînement du modèle sans revenus pour prédire les revenus :

Pour entraîner le modèle, nous avons utilisé des caractéristiques telles que Day, Event_coded, Cycle_de_rechargement, Action_Businnes, et Jour_ferie, en omettant temporairement la colonne Revenu. Nous avons appliqué l'algorithme de clustering KMeans avec 20 clusters pour segmenter les données en groupes similaires.

Chaque cluster a été analysé pour comprendre ses caractéristiques spécifiques et identifier des patterns récurrents. Pour représenter les revenus de chaque cluster, nous avons utilisé les centroïdes calculés par le modèle et avons visualisé ces centroïdes en utilisant le 75ème quantile (quantile 75%) des dates de chaque cluster pour une

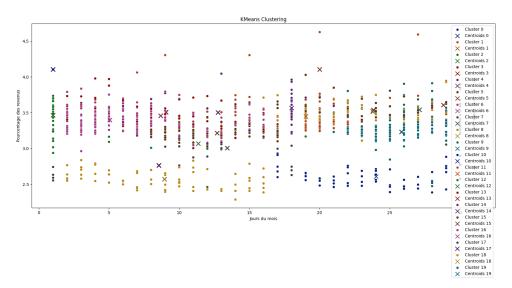


Fig. 3.2 Visualisation des clusters et de leurs centroids

représentation plus robuste des comportements observés. Cette approche a permis de mieux capturer les variations des données et d'éviter l'influence des valeurs extrêmes qui pourraient affecter la moyenne.

	count	mean	std	min	25%	50%	75%	max
Cluster								
0	65.0	3.317498	0.178731	2.8914	3.204400	3.30830	3.444200	3.7466
1	17.0	3.607165	0.184992	3.2931	3.516200	3.58730	3.685500	3.9563
2	10.0	3.020800	0.309977	2.4625	3.000300	3.13020	3.208700	3.3676
3	55.0	3.378285	0.167196	3.0325	3.269450	3.36920	3.469750	3.7665
4	30.0	3.462260	0.329680	2.6291	3.407575	3.50800	3.673775	3.9647
5	65.0	3.326165	0.218908	2.3866	3.236800	3.32980	3.444300	3.8332
6	25.0	3.325168	0.365448	2.5553	3.237200	3.44590	3.581900	3.7363
7	60.0	2.627990	0.206424	2.3878	2.508600	2.60225	2.710900	3.8501
8	52.0	3.533173	0.176895	3.1937	3.438250	3.51475	3.624825	4.0362
9	61.0	3.296908	0.201114	2.9659	3.200700	3.25440	3.365000	4.3066
10	5.0	3.374560	0.415607	3.1056	3.141000	3.24750	3.271200	4.1075
11	54.0	3.327515	0.198898	2.5477	3.245725	3.29170	3.428150	3.8045
12	58.0	3.527674	0.238964	3.0555	3.402350	3.48065	3.625750	4.6277
13	54.0	3.459893	0.184212	3.0495	3.341425	3.46245	3.544500	4.0653
14	63.0	3.481541	0.280728	2.3977	3.373800	3.49460	3.607400	4.3073
15	52.0	3.365842	0.170665	2.8497	3.280350	3.33395	3.445775	3.9063
16	59.0	3.311047	0.154307	3.0588	3.213450	3.26940	3.381900	3.7362
17	51.0	3.346037	0.242837	3.1017	3.215350	3.27750	3.382900	4.5953
18	59.0	2.599997	0.192431	2.2872	2.468300	2.56580	2.689700	3.5550
19	17.0	3.503300	0.251057	3.0745	3.345500	3.56400	3.620100	4.0486

Fig. 3.3 Les différentes distributions des revenus pour chaque cluster

3 Prédiction des revenus d'un nouveau mois sans données de revenus et calcul de l'erreur :

La dernière étape a consisté à utiliser les clusters pour prédire les revenus d'un mois sans données de revenus disponibles. Nous avons attribué un cluster à chaque jour du nouveau mois basé sur ses caractéristiques en utilisant le model KMeans ultérieurement sauvegardé, puis utilisé le 75ème quantile (quantile 75%) des revenus observés (Ultérieurement sauvegadé dans un fichier CSV) dans chaque cluster pour estimer les revenus journaliers. Ces estimations ont été compilées pour prédire le revenu total du mois.

Pour évaluer la précision de nos prédictions, nous avons comparé les revenus prévus aux données réelles une fois disponibles. Les erreurs ont été calculées en utilisant des métriques telles que l'erreur absolue moyenne (MAE). Cette analyse d'erreur nous a permis de mesurer la performance du modèle et d'identifier des opportunités d'ajustement pour affiner les prédictions futures.



Conclusion

Notre projet de développement d'un modèle prédictif chez Djezzy, basé sur le clustering pour l'estimation des revenus mensuels, a abouti à des résultats remarquables. L'approche innovante adoptée a permis de :

- 1. Segmenter efficacement les données grâce aux techniques de clustering avancées.
- 2. Identifier des tendances récurrentes cruciales pour l'analyse.
- 3. Générer des prévisions précises des revenus mensuels, même pour des périodes sans données disponibles.

L'efficacité du modèle est attestée par une erreur absolue moyenne (MAE) exceptionnellement basse de 0.05, soulignant sa fiabilité et sa précision. Cette performance valide la pertinence de notre approche et son potentiel pour guider les décisions stratégiques de l'entreprise.

Ce projet démontre non seulement l'efficacité du clustering dans la prédiction des revenus, mais illustre également l'impact concret des techniques d'analyse avancées dans un contexte commercial. Les résultats obtenus constituent une base solide pour l'optimisation future des stratégies commerciales de Djezzy, ouvrant la voie à une gestion plus informée et proactive de l'entreprise.