

An examination of the “The movement of people” using the UN Data sources

Rodney Wardle SBS23057

Capstone Project Final Poster For Strategic Thinking

Introduction

This capstone project was to look at UN data for asylum seekers application globally and see how a machine learn model can be used successfully on the given data from the UN and make predictions on asylum applications now and in to the future, It was supplied from the official UN data website. This data is based off global figures for asylum seeker applications and to create machine learning models to predict future applications.

This capstone project followed the CRISP DM Methodology

Business Understanding → Data Understanding → Data Preparation → Modelling → Evaluation → Deployment

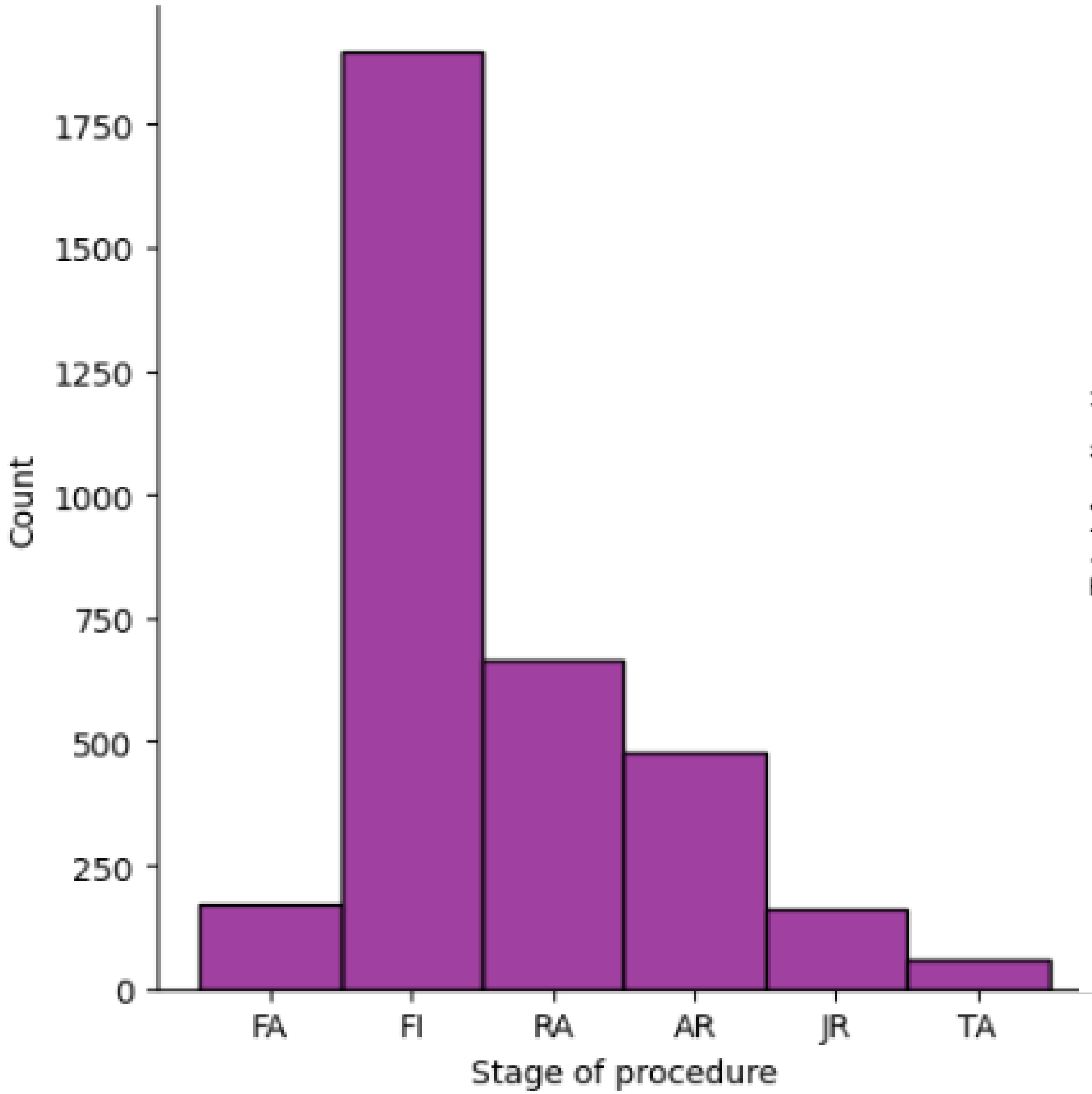
5 Main Objectives

- 1.How can data science be used to analyze the growing number of asylum seekers around the world.
- 2.Examination of available data may help to predict the future applications for asylum seekers across the world
- 3.To develop a machine learning model to estimate the number of asylum applications. Following the CRISP-DM Project management framework for data analysis.
- 4.Compare the estimates of applications both supervised and unsupervised and a description of exactly what this entails.
- 5.To take a deeper look into the global figures to allow for some examination of the movement of peoples within Europe, America, Asia and Oceania (Australia).

Technologies Used

- Libraries Used: pandas,seaborn, numpy,sklearn,matplotlib etc.
- For hyperparameter tuning Grid Search CV was used.
- Machine Learning Algorithm used: Decision Tree, K Nearest Neighbor and K Means
- Label Encoder was used to deal with the categorical data

Visualizations



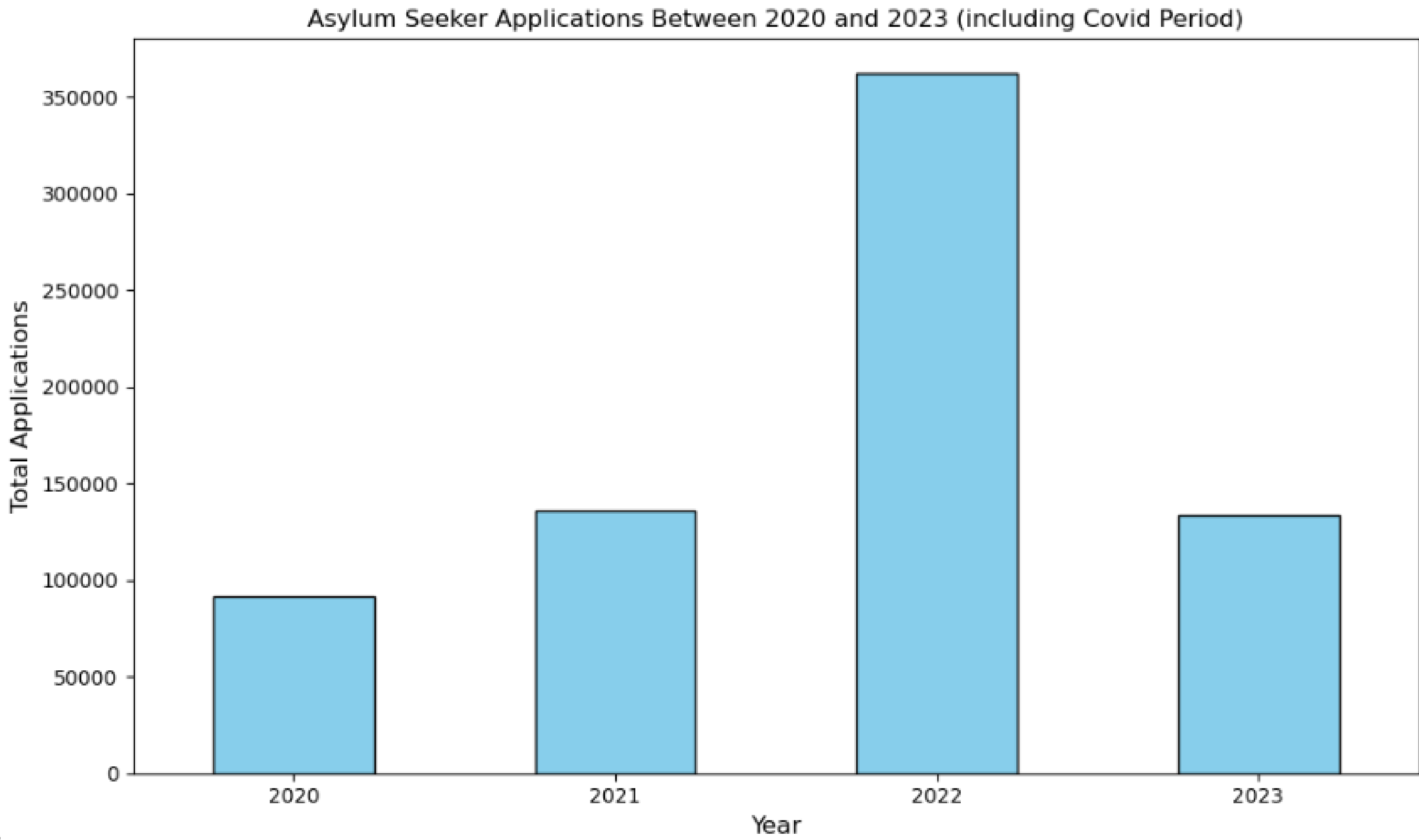
Model Evaluation Key Metrics

Precision: this focuses on the proportion of the true positive predictions amongst all predictions

Recall: This focuses on the proportion of true positives among all the actual positive ,

F1 Score: This combined the precision and recall into a single metric.

Support: This focuses on the number of actual occurrences of the class in the dataset.



Model Evaluation

Model	Precision	Recall	F1 Score
Decision Tree	0.24	0.93	0.39
KNN	0.21	0.87	0.46

Best support for the decision was at 246

K Means clustering best model results:

Best hyperparameters found by GridSearchCV: {'init': 'random', 'max_iter': 100, 'n_clusters': 8, 'n_init': 10}

Conclusion

We have been able to show successfully that a machine learning algorithm can be used in the prediction of the amount of asylum seekers applications and to prepare for the future needs for the asylum seekers globally as a whole.

More work is required on the algorithms to get the accuracy up higher into 90% +.

More finetuning with the hyperparameter is needed to get the best results possible.

Challenges Encountered

- A full deep clean of the data was needed. Converting from categorical to numerical This took the longest the data understanding and data preparation.
- It was extremely hard to interpreted the results from these models. The hardest model was the Decision Tree Classifier.
- In future work I need to look at the different features in deeper detail, I just scrapped the surface on this dataset, it is quite complex in nature.

References

unstats.un.org. (n.d.). UNSD — Demographic and Social Statistics. [online] Available at: <https://unstats.un.org/unsd/demographic-social/concerns/migration/>.

Smart Vision Europe (2017). Building and Applying Predictive Models in IBM SPSS Modeler training webinar. [online] Smart Vision - Europe. Available at: <https://www.sv-europe.com/crisp-dm-methodology/>.

IBM (2023). What is a Decision Tree | IBM. [online] [www.ibm.com](https://www.ibm.com/topics/decision-trees). Available at: <https://www.ibm.com/topics/decision-trees>.

Team, G.L. (2023). Label Encoding in Python - 2023. [online] Great Learning Blog: Free Resources what Matters to shape your Career! Available at: <https://www.mygreatlearning.com/blog/label-encoding-in-python/#:~:text=Label%20encoding%20is%20a%20technique>.

Kaushik, S. (2019). An Introduction to Clustering & different methods of clustering. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>.