**Abstract**

**Index:**

1. **Introduction**
2. **Data Preparation and Visualization**
   1. EDA
   2. Method justification
   3. Visualisations
   4. Tuft Principles
3. **Statistical Analysis**
   1. Dataset summary
   2. Binomial and Poisson Distribution
   3. Normal Distribution
   4. Importance of Binomial and Poisson and variable justification. Discrete variables in a Normal Distribution
4. **ML**
   1. Project management framework CRISP-DM, KDD and SEMMA. Supervised, unsupervised, or semi-supervised machine learning technique.
   2. ML models
   3. Results of ML models selected
   4. Similarities and differences of ML models selected
5. **Programming**
   1. Programming paradigms used
6. **Conclusion**
7. **References**
8. **Appendix**

**1. Introduction**

The aim of this study is to highlight the population changes for the Republic of Ireland. We will not go beyond 1923, however we do have until 2023. In my opinion this is enough time to get a picture of the current demographics. We will get into detail in each of the sections.

**2. Data preparation and visualization**

2.1. EDA

2.2. Method justification

2.3. Visualizations

2.4. Tuft Principles

The aim of this

**4. ML**

4.1. Project management framework CRISP-DM, KDD and SEMMA.

As a project management framework, we are going to discuss each of them in detail. We have CRISP-DM (Cross Industry Standard Process project), KDD (Knowledge Discovery in Databases) and SEMMA (Sample, Explore, Modify, Model, Assess) models.

CRISP-DM has six phases, business/research understanding, data understanding, data preparation, modeling phase, evaluation, and deployment. A real-world example for this framework would be a hospital predicting the length of stay for its patients, based on a series of indicators. This method was conducted using Average Prediction, Multiple Regression, Decision Tree, Artificial Neural Network ensemble, Support Vector Machine and Random Forest. The implementation of this model lead to building successful predictions for the hospitals (Caetano et al., 2015).

A diagram of a data flow

Description automatically generated

Figure : CRISP-DM Flow

KDD in this framework we retrieve and analise the data stored in databases, the entire process consists of seven steps, data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge representation. A good example of this method is the use of machine learning in the prediction of sports results. A recent study has been conducted to predict football matches (Głowania et al., 2023).

**A diagram of data processing

Description automatically generated**

Figure : KDD Flow

SEMMA entails five steps, sample, explore, modify, model and assess. This method has been used for Crime Prediction and conducted using article neural networks machine learning concept (Forradellas et al., 2020).

A blue rectangle with white text

Description automatically generated

Figure: SEMMA flow

In PEA20.20231004T131025.csv we have labelled data that is why I have selected supervised machine learning techniques, we can train the model in regard to migration flows depending on people economic status, sex and inward or outward flow.

4.2. ML models

1. Programming
   1. Programming paradigms used