PREDICTING DIABETES READMISSION

Li-Tin (Lily) Chen

CAPSTONE PROJECT

Bio

Li-Tin (Lily) Chen

Data Scientist from Institute of Data

Bachelor of Midwifery

Diabetes in Australia

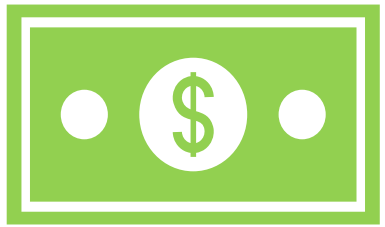
In 2016-2017

1.2 MILLION

diabetes-related hospitalisations

10% of all hospitalisations

Annual Health Expenditure:



$6 BILLION

13% readmitted within 30 days

What can be done?

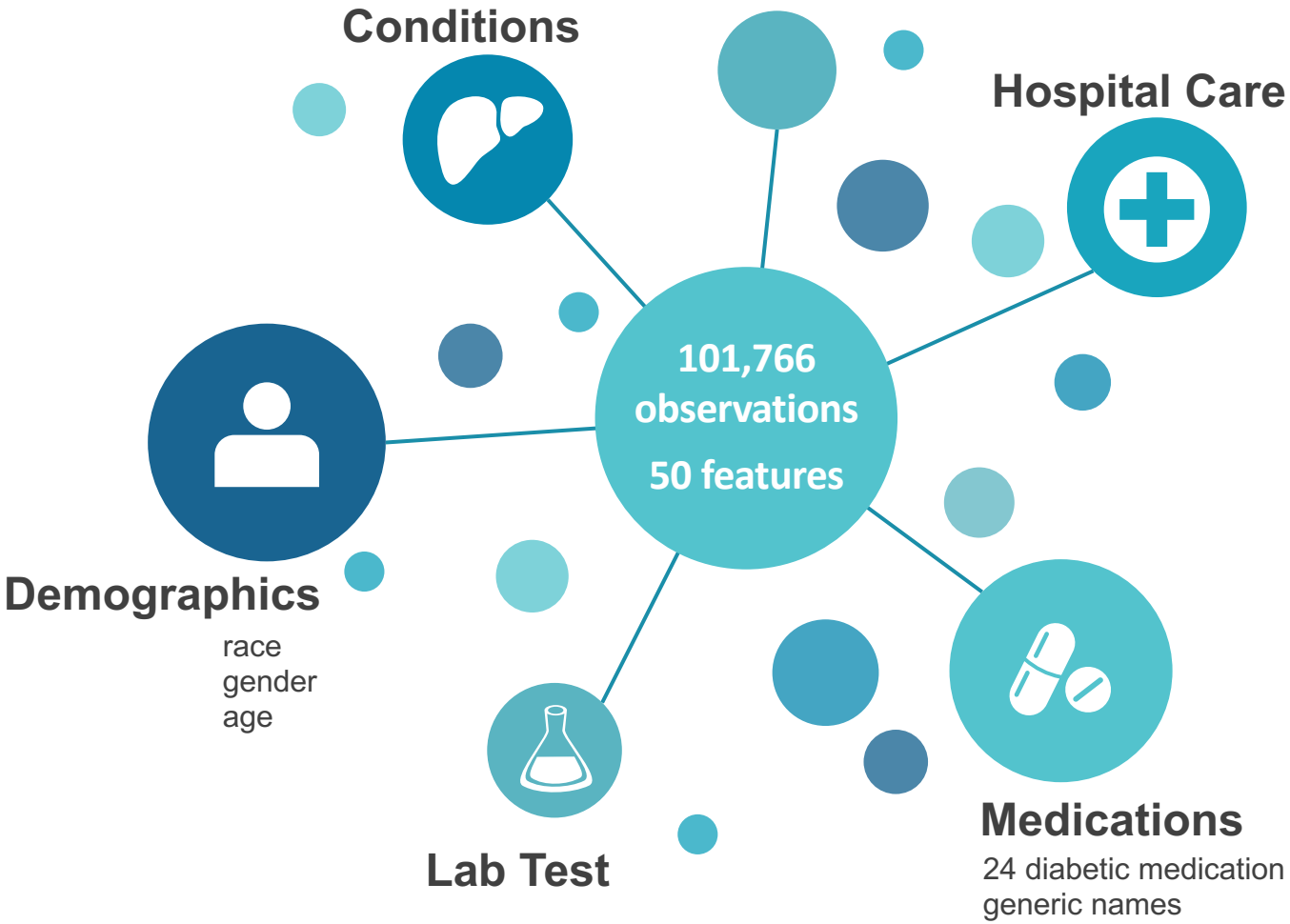
How to lower hospital readmission rate for diabetic patients to minimise unnecessary spending?

Average cost per hospital admission: $7656
Current readmission rate: 13%

10% readmission ↓ ↑ **$230** million saved

1. What factors show strong correlation with readmission rate?
2. What model can we use to accurately predict future readmission based on existing data?

Dataset Description



Source: UCI repository

Collection Timeframe: 1999-2008

- Contain missing values
- Incorrect data type

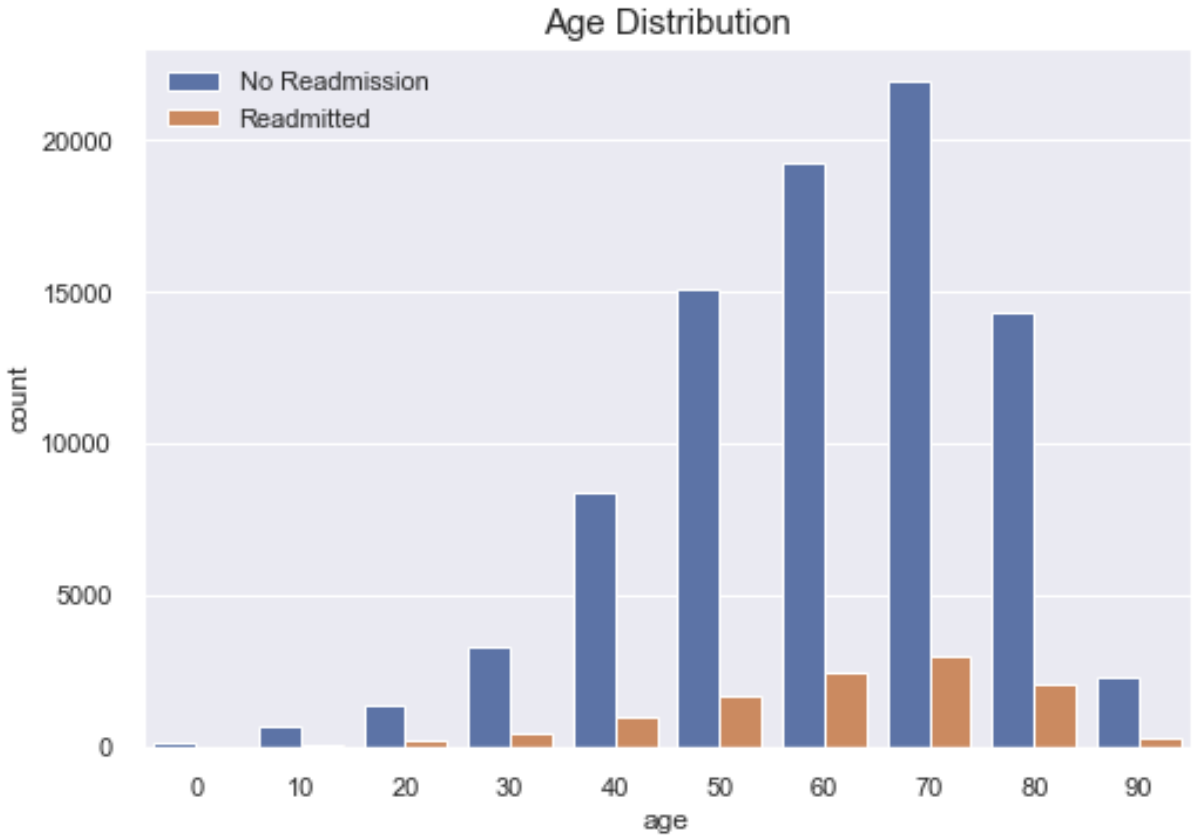
Final Cleaned Dataset:

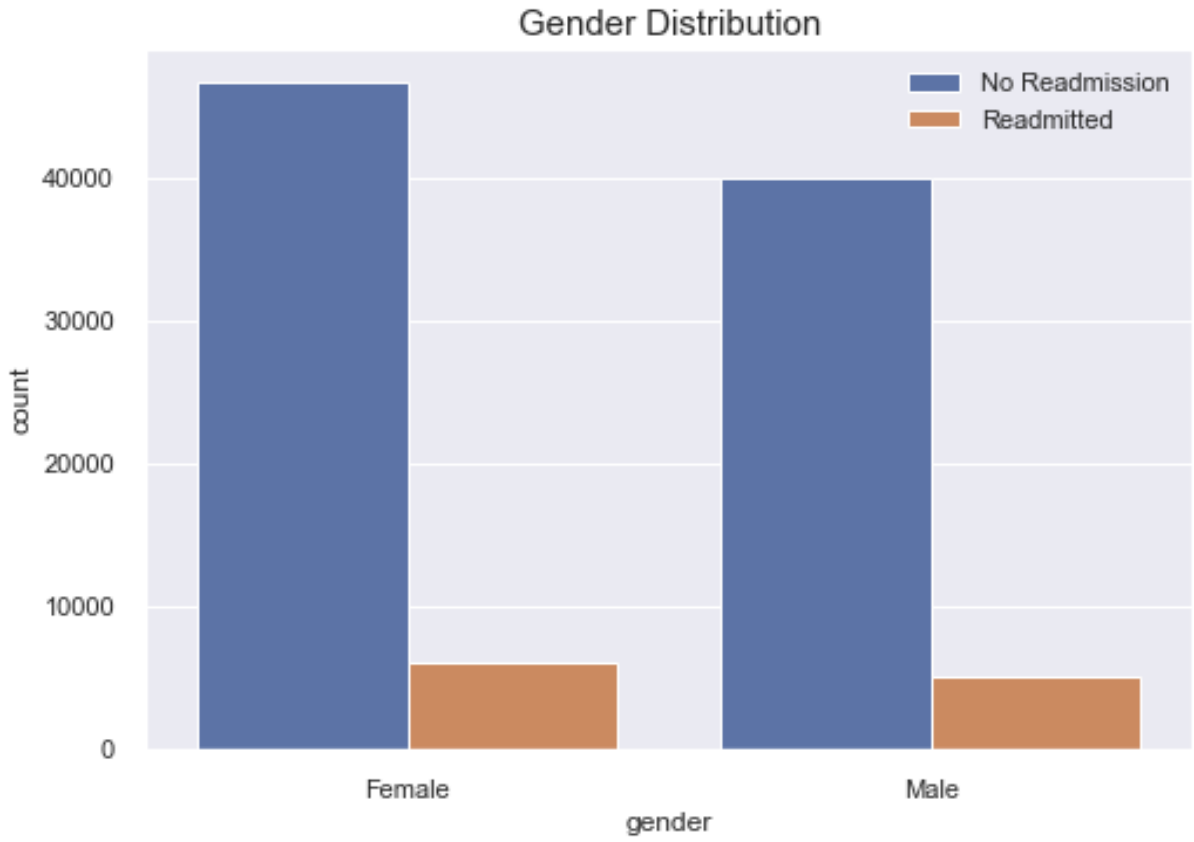
- 97874 rows x 15 columns
- No missing values
- All data of numeric type

Outcome variable: readmitted

- 0 = No readmission
- 1 = Readmitted

Data Exploration





Imbalanced outcome distribution:

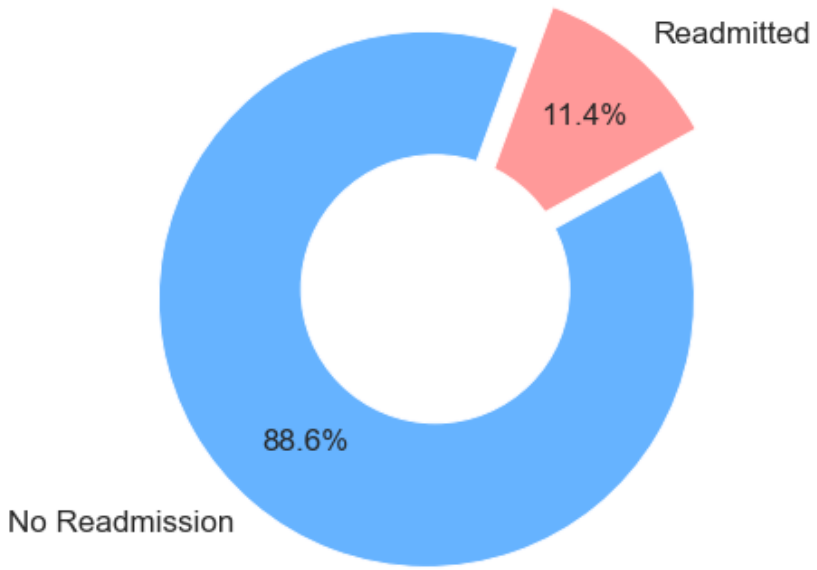
- **High in non-readmission (0)**
- **Low in readmitted (1)**

Similar features distributions between the two outcome classes – low correlation

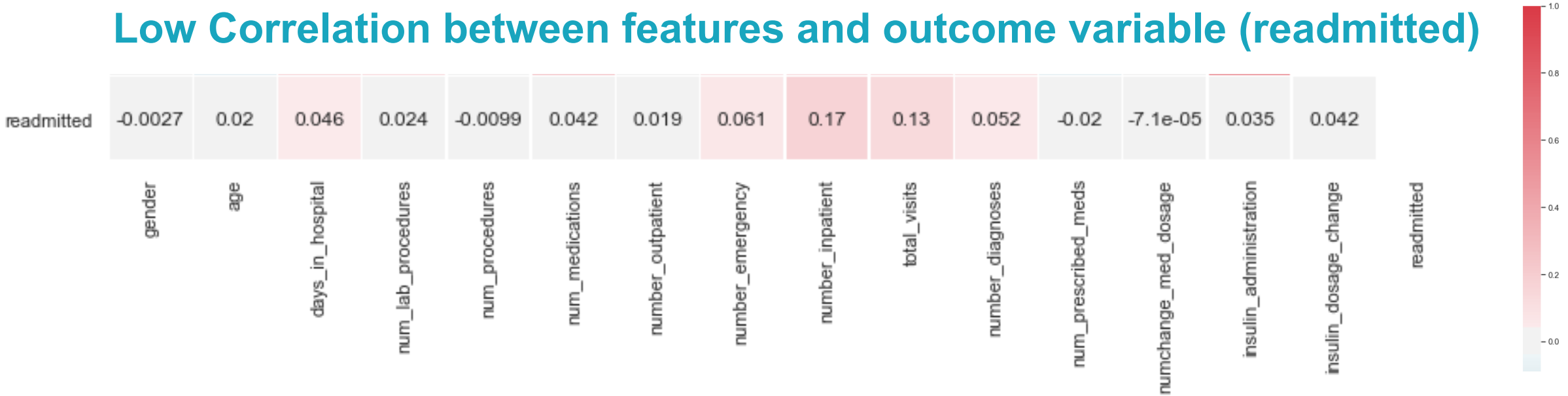
EDA Findings

Imbalanced Dataset

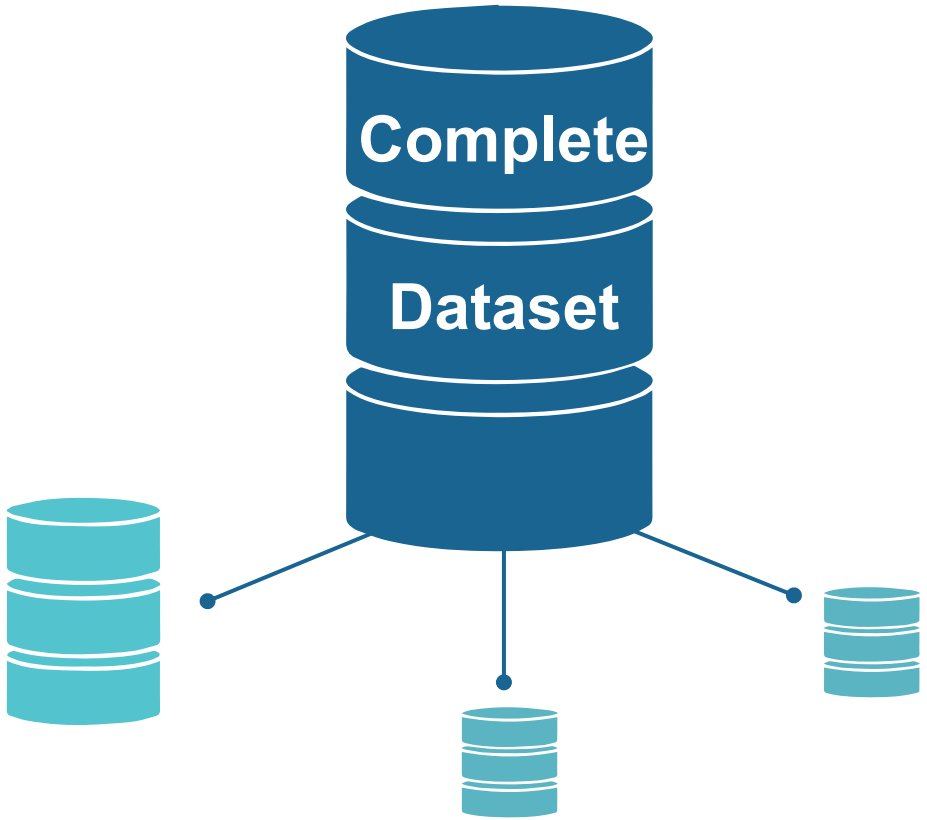
	Outcome	Number of Patients
0	No Readmission	86705
1	Readmitted	11169



Low Correlation between features and outcome variable (readmitted)



Data Allocation



Train Set
60%
Model training purpose

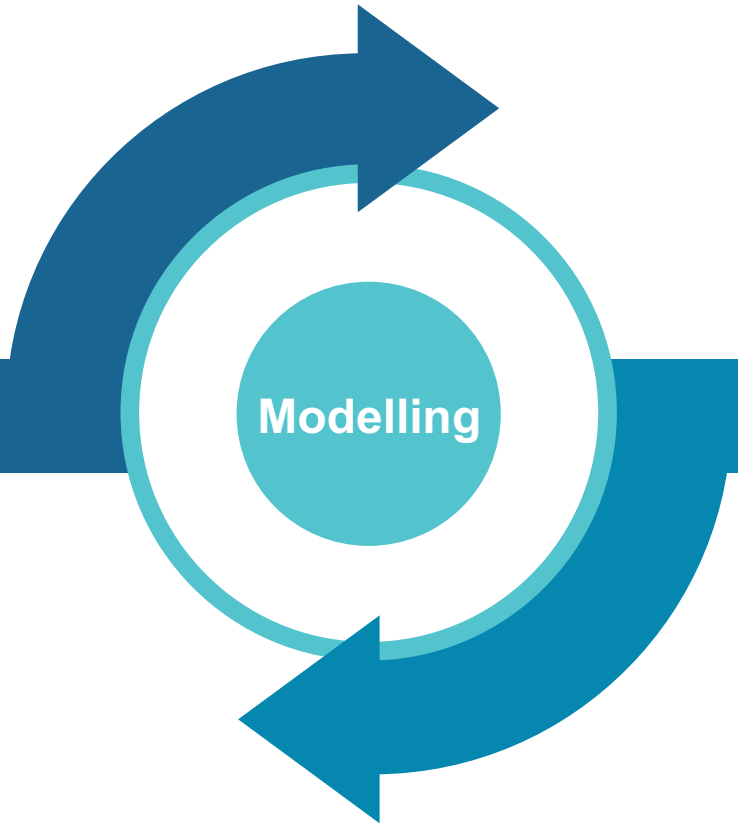
Test Set
20%
Model configuration and hyperparameters tuning

Validation Set
20%
'Unseen' Data
Unbiased evaluation on the final model

Data Processing

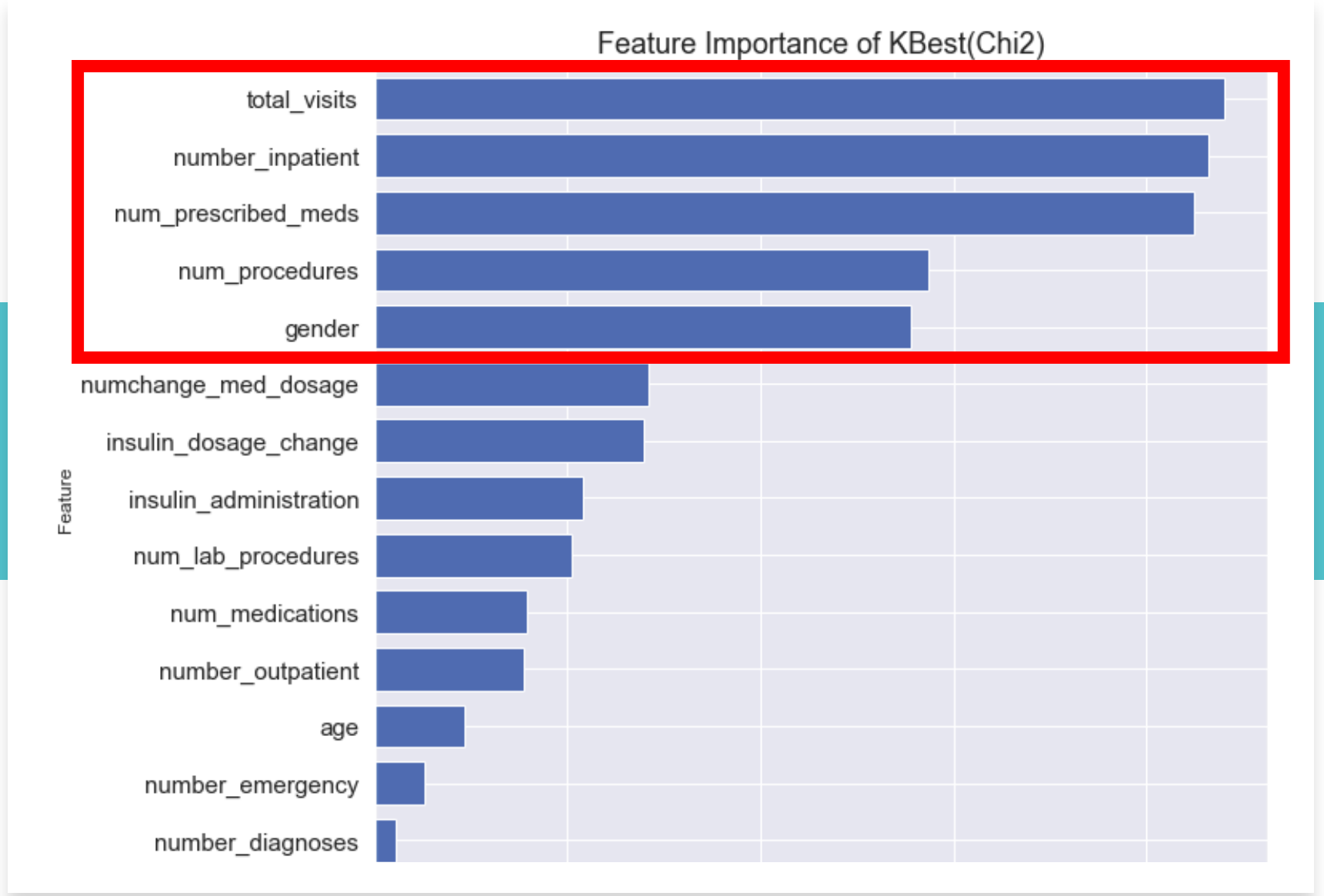
Imbalanced data leads to bias in favour of the majority class (No readmission)

Data Balancing



Feature Selection

Identify strong predictors of the outcome



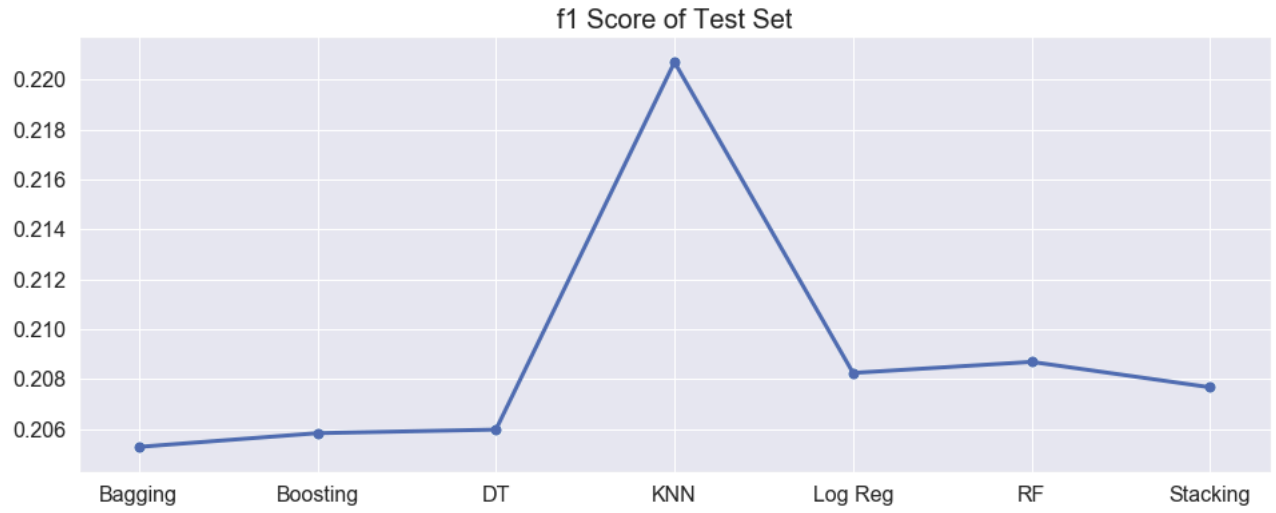
Top 5 Features

Top Contributing Features

Models Evaluation

- 7 Models
- Randomised search with 5-fold cross validation for finding best hyperparameters combinations.

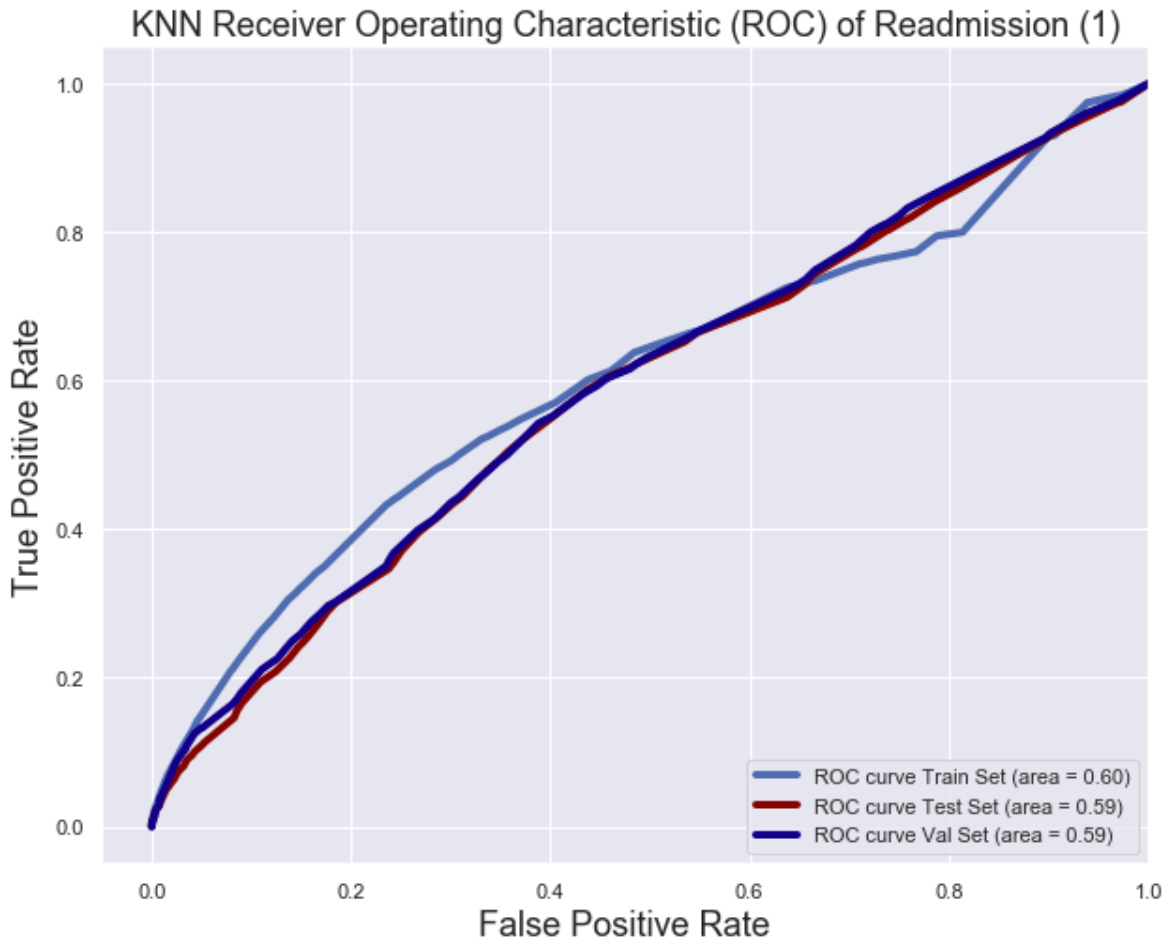
	Model	Recall	Precision	f1 Score
0	Log Reg	0.465	0.134	0.208
1	KNN	0.356	0.160	0.221
2	DT	0.497	0.130	0.206
3	RF	0.504	0.132	0.209
4	Bagging	0.495	0.130	0.205
5	Boosting	0.498	0.130	0.206
6	Stacking	0.480	0.132	0.208



Final Model - KNN

Recall: 36% of actual readmissions were labelled correctly
Precision: 16% of all predicted readmissions were actual readmissions
F1 score of 0.22 – harmonic mean between precision and recall

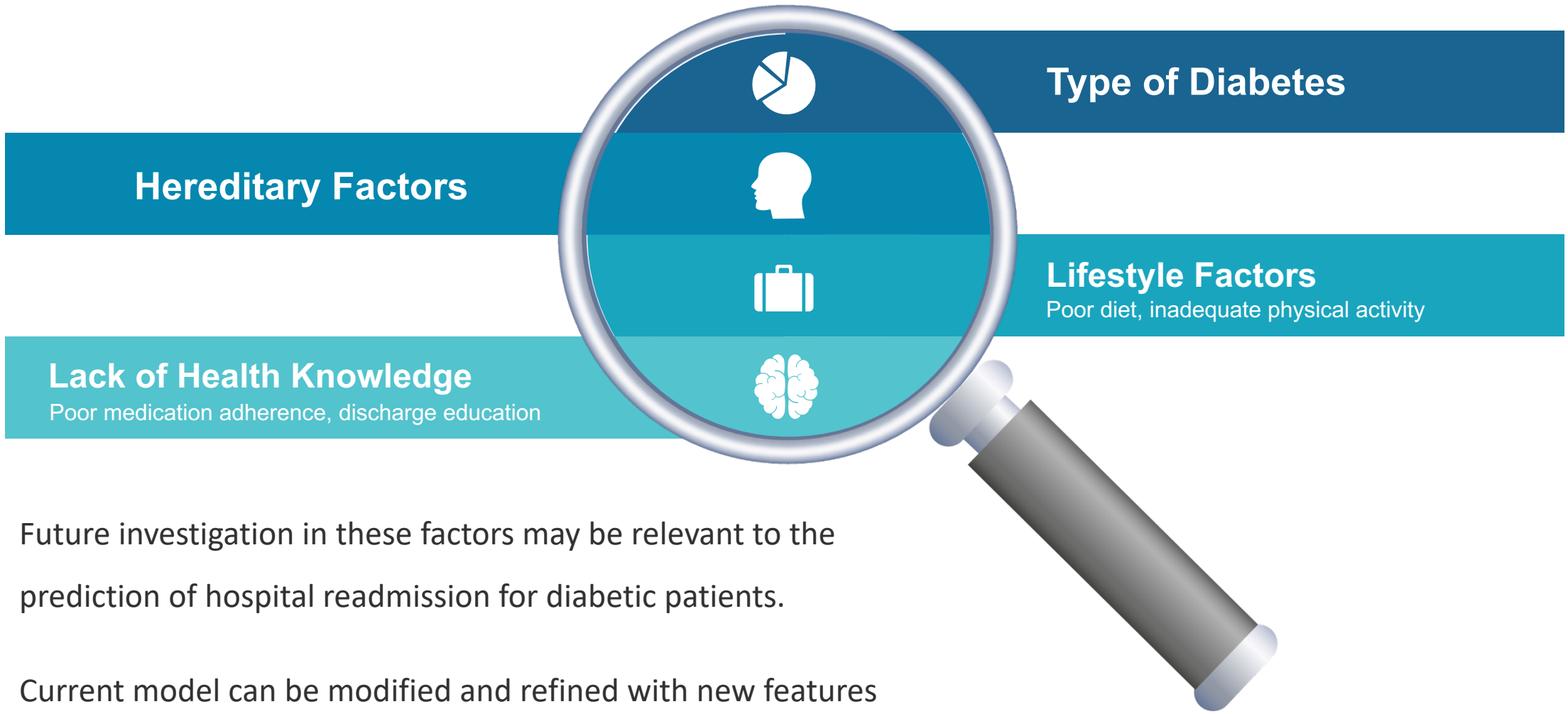
KNN Performance on ‘Unseen’ Data



	Train Set	Test Set	Val Set
Accuracy	0.599	0.711	0.714
AUC	0.599	0.556	0.560
Recall	0.437	0.356	0.359
Precision	0.646	0.160	0.166
f1 Score	0.521	0.221	0.227

Main Finding: Factors we thought might be useful (prior to data analysis) is found to give low predictive values in predicting the readmission.

Recommendations



Future investigation in these factors may be relevant to the prediction of hospital readmission for diabetic patients.

Current model can be modified and refined with new features

References

Supporting Documentation on GitHub Repository:

- Exploratory Data Analysis: Lily-Chen/capstone/code/Lily Capstone Project Diabetes Hospital Readmission EDA.ipynb
- Modelling: Lily-Chen/capstone/code/Lily Capstone Project Diabetes Hospital Readmission Modelling FINAL.ipynb

Data Source: https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008#

Websites:

- https://static.diabetesaustralia.com.au/s/fileassets/diabetes-australia/e7282521-472b-4313-b18e-be84c3d5d907.pdf
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4966497/#:~:text=Poor%20medication%20adherence%20in%20T2D,and%20managing%20complications%20of%20diabetes.
- https://towardsdatascience.com/hyper-parameter-tuning-and-model-selection-like-a-movie-star-a884b8ee8d68
- https://www.aihw.gov.au/reports/diabetes/diabetes/contents/what-is-diabetes
- https://www.sciencedirect.com/science/article/pii/S1059131117302455
- https://link.springer.com/article/10.1186/s12913-018-3723-4
- https://lukesingham.com/whos-going-to-leave-next/
- ttps://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226
- https://link.springer.com/article/10.1007/s11892-018-0989-1


THANK YOU