An Analysis of the Location Data of Hospitals in Metro Manila

GABRIEL URETA

HTTPS://GITHUB.COM/GPURETA/COURSERA_CAPSTONE.GIT

Government vs Private Hospitals

When there is an opportunity to open a business, existing infrastructure affects the type of costumers present

In the Philippines, government hospitals are more affordable than private hospitals but less accessible due to long lines and schedules

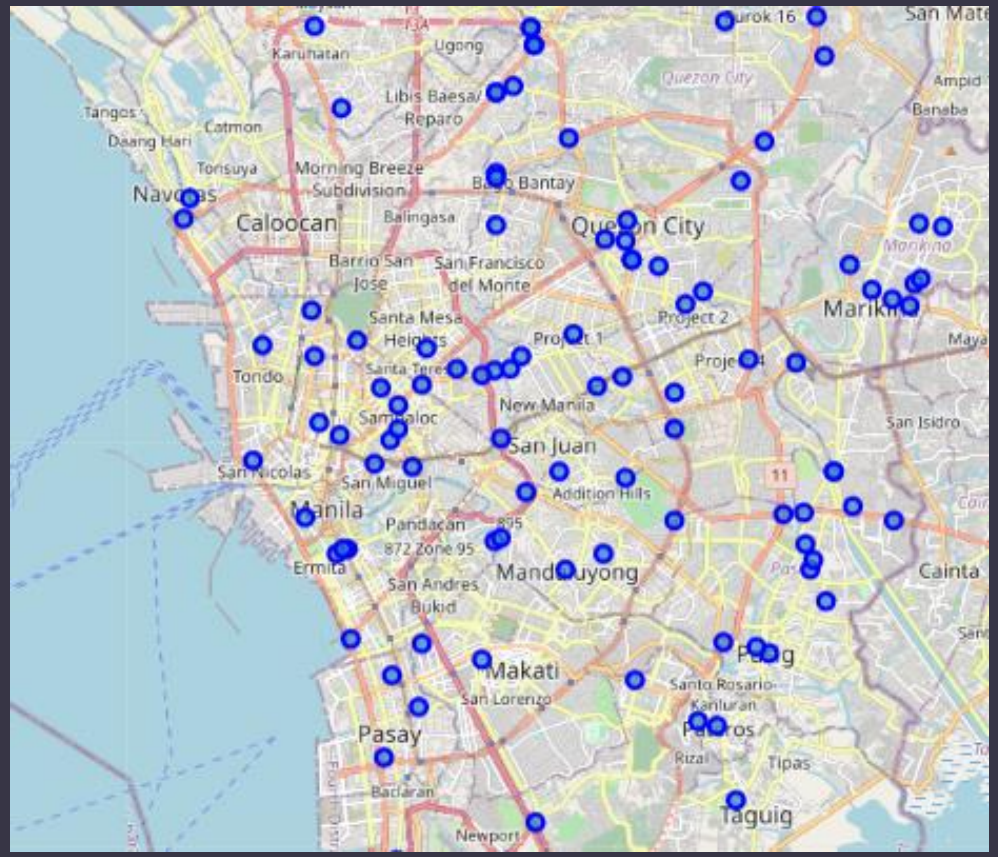
Private hospitals cater to higher income patients by providing more comforts and amenities at higher price

It would be worthwhile to learn how the type of hospital affects the type of businesses and venues nearby due to the difference of costumer demographics

Business Problem

1. What are the differences between the type of businesses/venues that surround government and private hospitals?
2. Can we predict if a hospital is government or private based on the venues that surround it?

Description of the Data Set



Hospitals considered are located in the National Capital Region of the Philippines (Metro Manila)

Hospital data was taken from the Philippine Department of Health website

Geopy was used to get the specific latitude and longitude for each hospital

Foursquare API was used to gather total venues at walking distance from the hospital (500 meters)

Exploratory Analysis

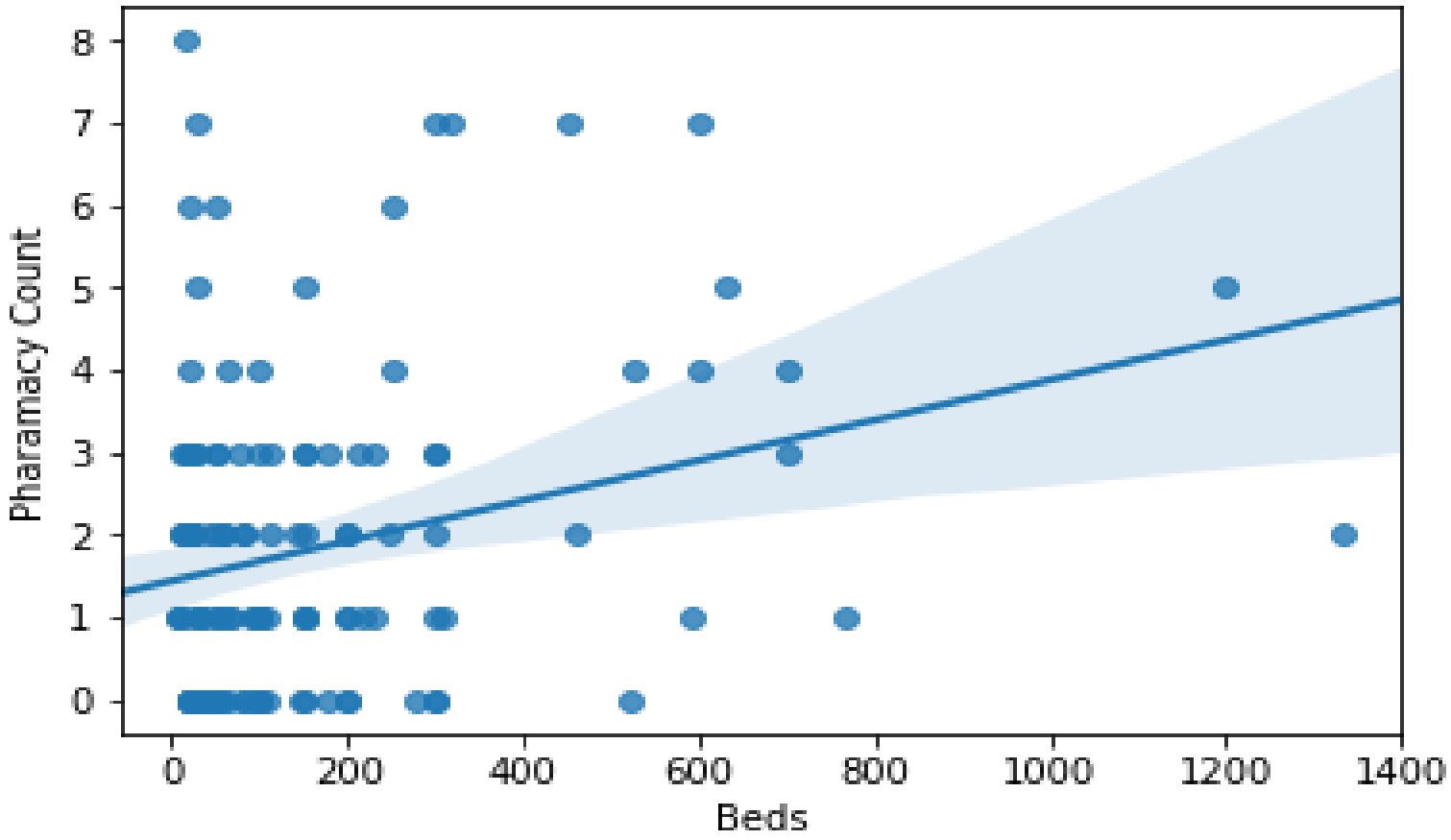
The effect of the size of the hospital on the surrounding venues needs to be taken into account

A larger hospital might cause a greater number of venues due to a higher potential customer traffic

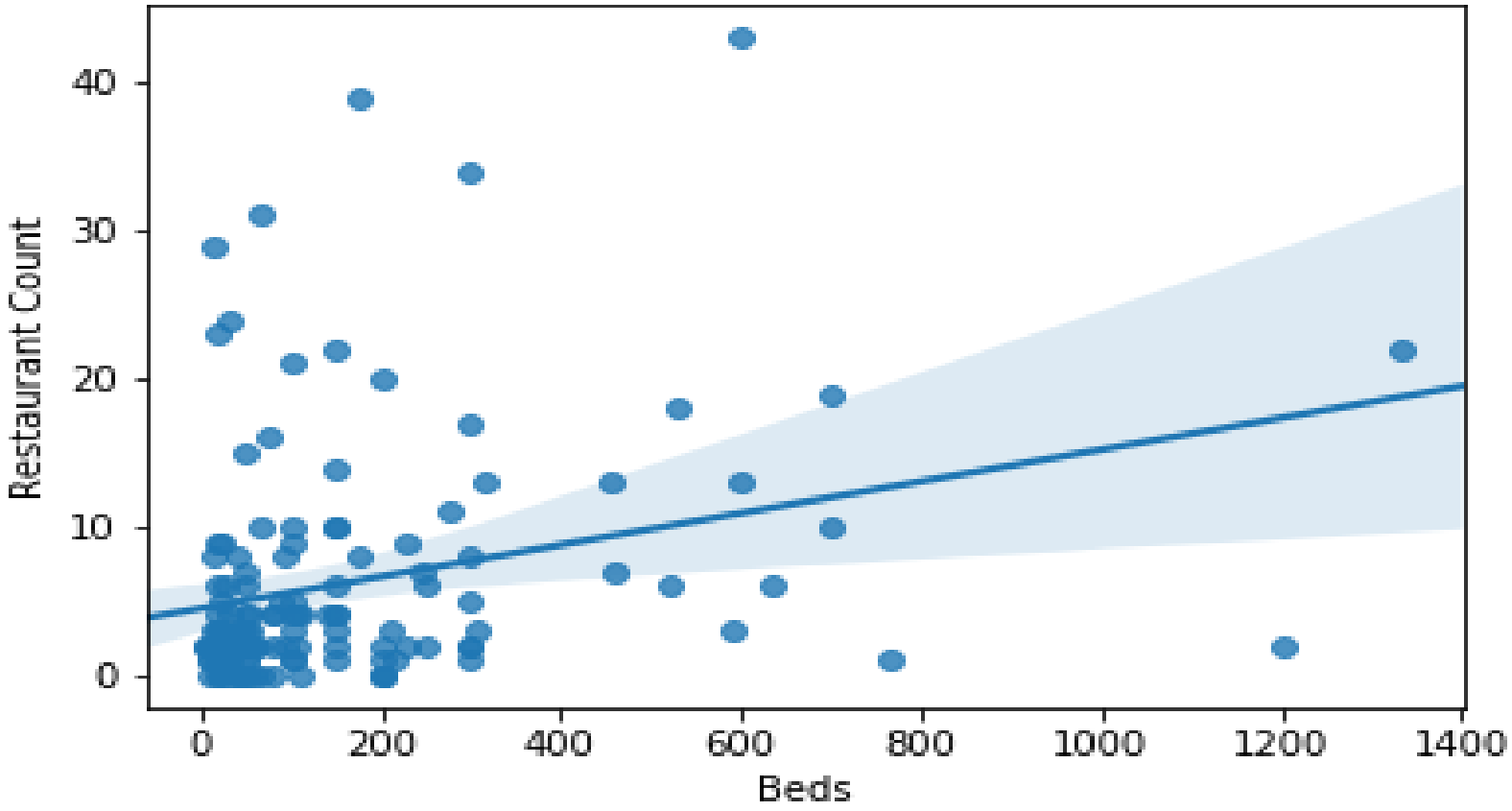
The data used to represent the size was the hospital bed capacity

The number of restaurants and pharmacies located at a walking distance (500 meters) from each hospital were then compared to its bed capacity.

No linear correlation between the bed capacity of the hospital and the number of pharmacies



No linear correlation between the bed capacity of the hospital and the number of restaurants



The Pearson Correlation Coefficient measures linear correlation

A value closer to +1 indicates a positive linear correlation

A value closer to −1 indicates negative linear correlation.

A value near zero indicates zero correlation.

The correlation coefficient is close to zero which implies that there is no linear correlation between the bed capacity of the hospital any of the venues

	Correlation	P-value
Beds vs Restaurants	0.2727	0.0015
Beds vs Pharmacies	0.2705	0.0016

Is there a difference in the type of businesses between Government and Private hospitals?

From the 133 hospitals in the data, there were 2797 gathered venues with 234 unique categories

Analysis of variance (ANOVA) was performed on the data to compare the venues of Government and Private hospitals

ANOVA allows us to see the categories of venues that are significantly different

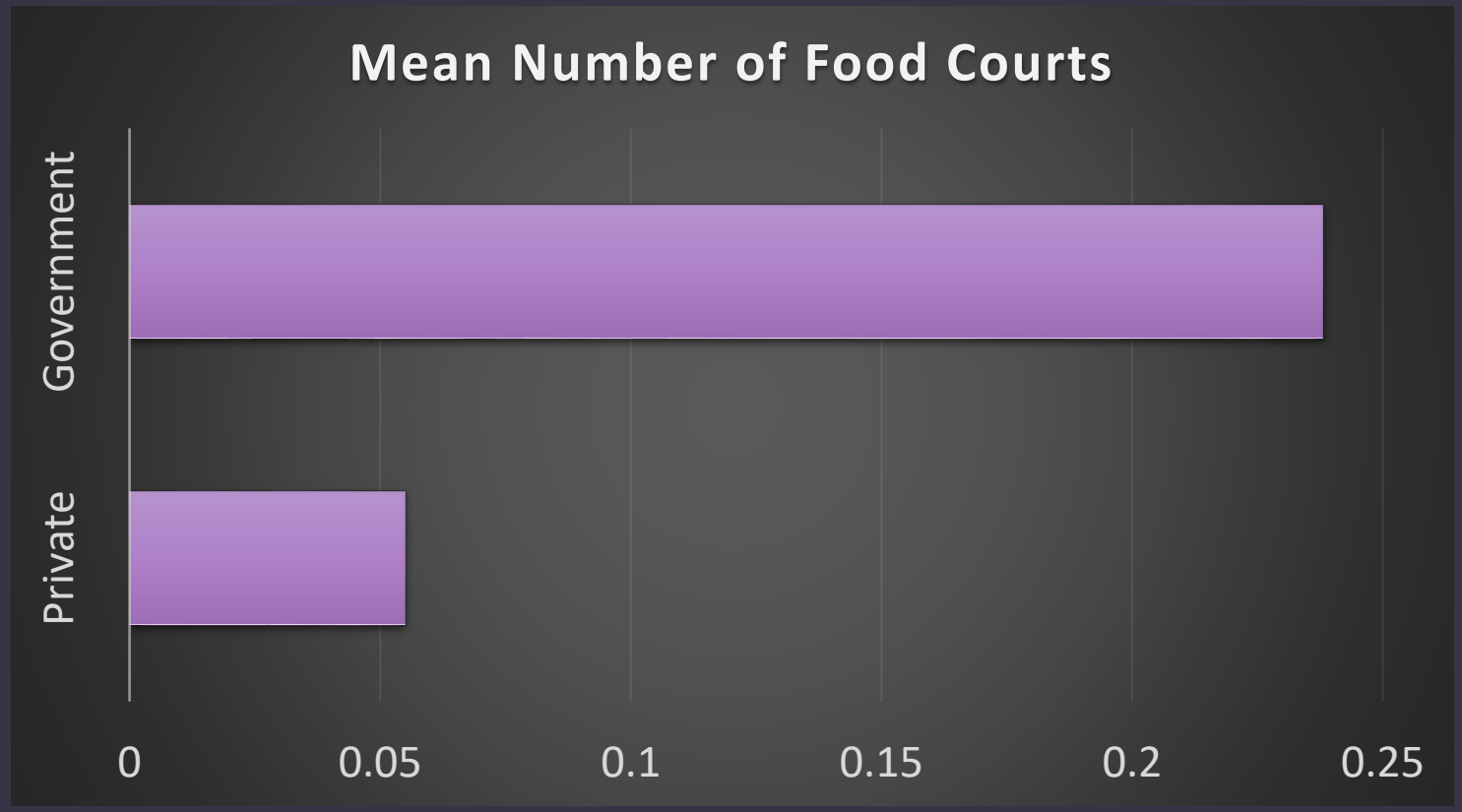
Significant Venues

Venues	P-Value
Chocolate Shop	0.036153
Food Court	0.003575
Fountain	0.036153
Indian Restaurant	0.036153
Museum	0.047205
Optical Shop	0.009674
Scenic Lookout	0.036153
Chocolate Shop	0.036153

A P-value with less than 0.05 suggests that the class (government/private) of the hospital affects the number of the specific venue category

The class of the hospital affects the number of chocolate shops, food courts, Indian restaurants, and optical shops.

There is moderate certainty that there are less food courts in private hospitals than in government hospitals



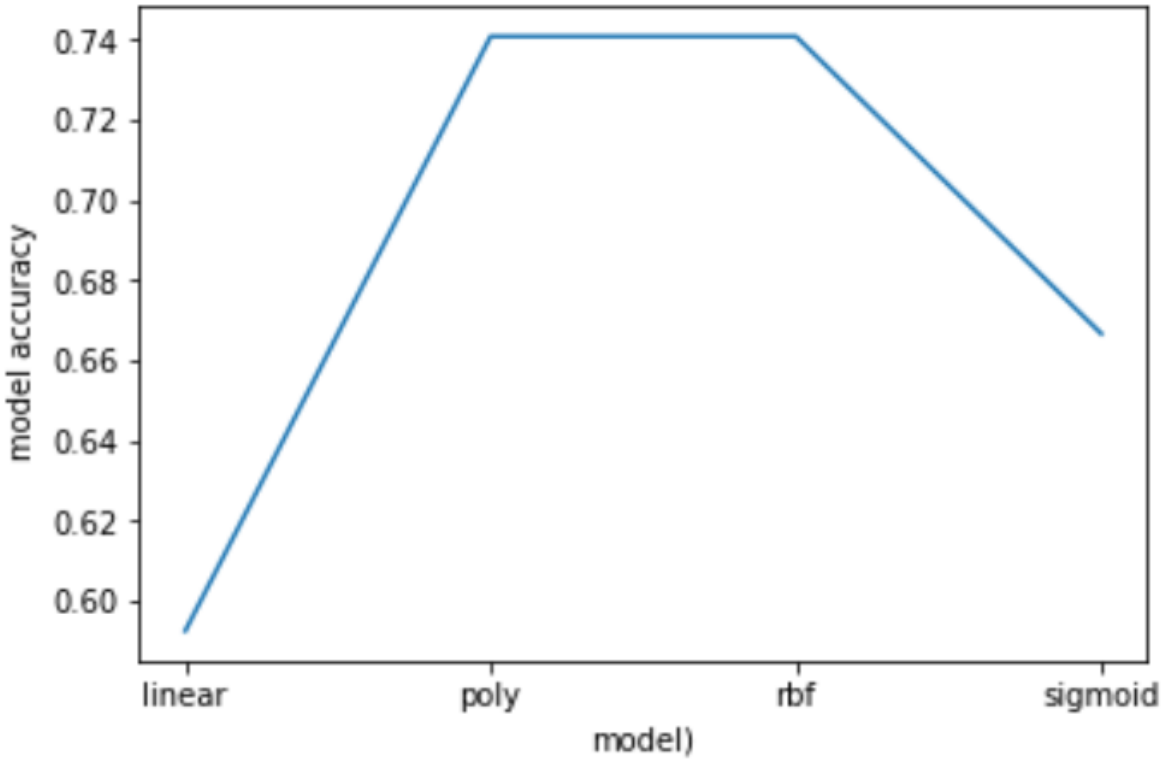
Predicting the Hospital Class

Machine learning classification models were used to see if they can predict the class of the hospitals based on the venues that exist around it

Data was split into 80% training data and 20% testing data to check the out-of-sample accuracy

Each model was optimized using the training data against the testing data, and then confirmed using the whole data against the testing data

Support Vector Machine

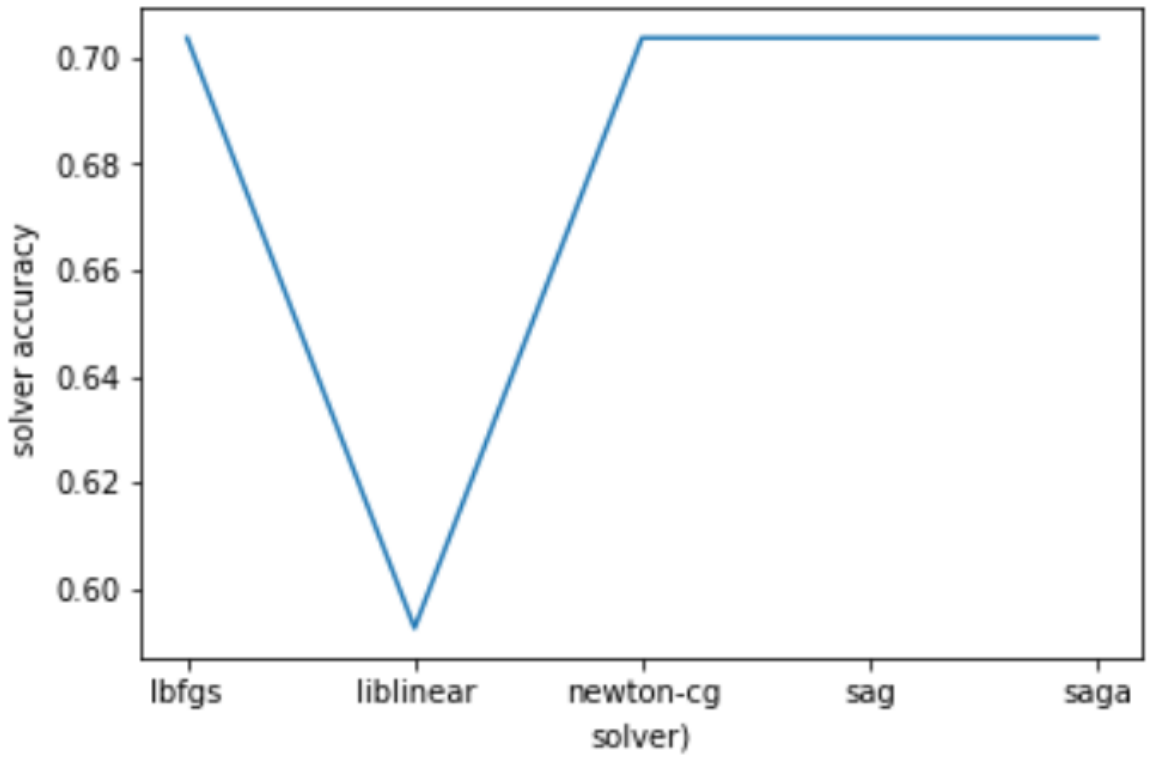


Model Performance:

Best models: Polynomial, RBF

Best accuracy: 0.74

Logistic Regression



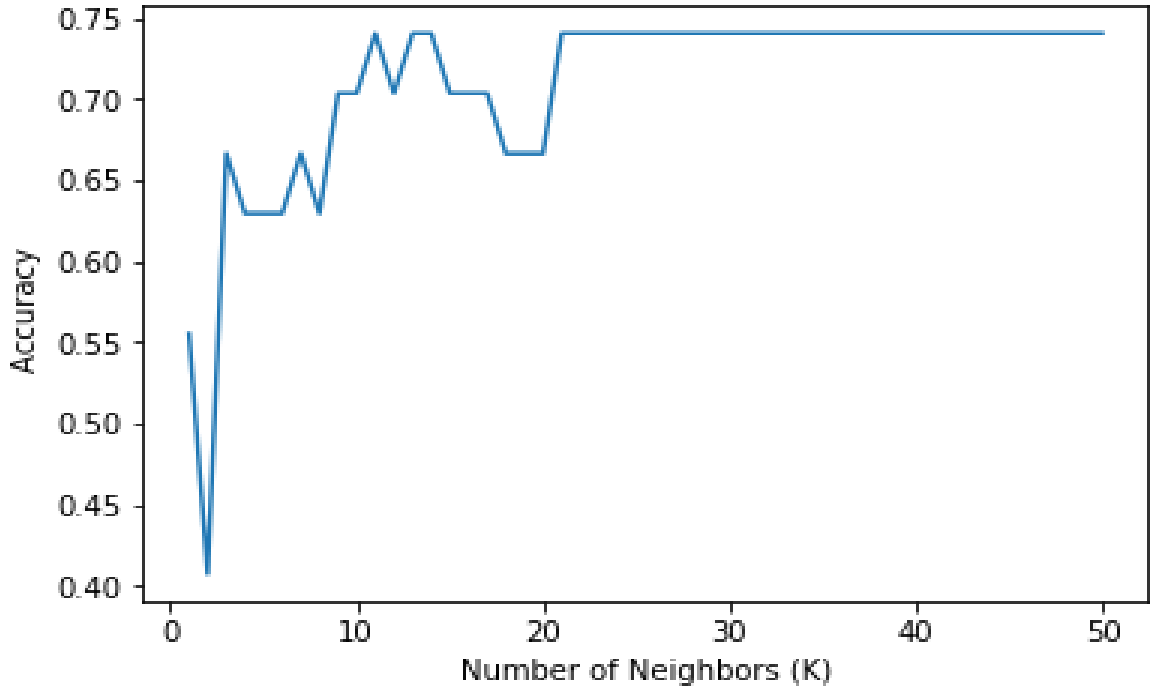
Model Performance:

Best models: Newton-CG

Best accuracy: 0.74

Log loss: 0.5

K Nearest Neighbours (KNN)



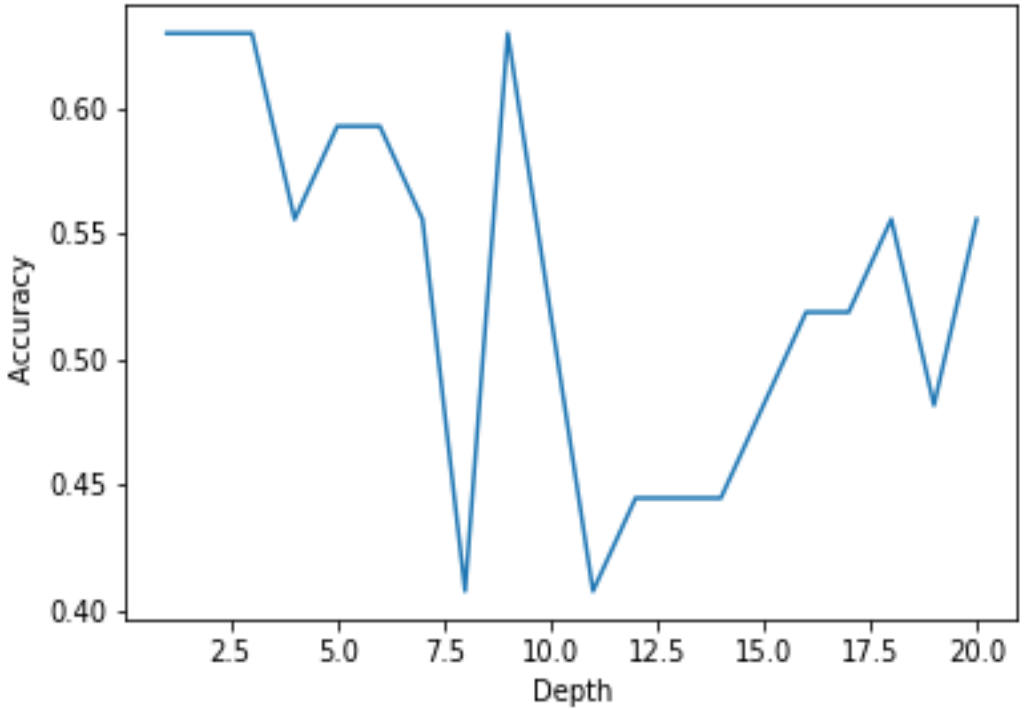
	precision	recall	f1-score	support
Government	0.50	0.14	0.22	7
Private	0.76	0.95	0.84	20
micro avg	0.74	0.74	0.74	27
macro avg	0.63	0.55	0.53	27
weighted avg	0.69	0.74	0.68	27

Model Performance:

Best models: k = 11, k =22

Best accuracy: 0.78

Decision Tree



	precision	recall	f1-score	support
Government	0.56	0.71	0.63	7
Private	0.89	0.80	0.84	20
micro avg	0.78	0.78	0.78	27
macro avg	0.72	0.76	0.73	27
weighted avg	0.80	0.78	0.79	27

Model Performance:

Best models: depth = 1, 9

Best accuracy: 0.78

Conclusions:

Bed capacity had no linear correlation on the number of businesses surrounding the hospital. Could be due to:

- Number of patients visiting the hospital is not proportional to hospital bed capacity
- Location demographics can affect the type of businesses present
- Hospital traffic not a factor in businesses

Predicting if a hospital is government or private based on nearby venues is possible:

- Decisiontree and KNN models had highest accuracy, but Decisiontree did not suffer low recall rates unlike KNN