D2i GEELONG CITY

Product Owner Presentation

City of Greater Geelong Analytical Dashboard

PROJECT OVERVIEW

Machine Learning and Data Analytics Capstone Project – School of IT, Deakin University

Focused on solving real world problem and communicating data artefacts related to the City of Geelong's Smart City initiative.

Working in close collaboration with Smart City Team, City of Greater Geelong Council

TEAM STRUCTURE

Team Member	Junior/Senior	Role
Srija	Senior	Data Analyst
Paul	Senior	ML Engineer
Vinayaka	Senior	ML Engineer
Omer	Senior	ML Engineer
Ngoc	Junior	ML Engineer
Shibendu	Junior	Data Analyst
Xiaoran	Junior	Full Stack Developer
Rishabh	Junior	Full Stack Developer
Yijia	Senior	Full Stack Developer
Anand	Senior	Full Stack Developer

SQUAD STRATEGY

Machine Learning

Research

Data Analytics and Modelling

Visualization

Web Development

ML model deployment

Designing UI

Content delivery

Agile Methodology

Trello Board

Cross functional team

Scrum meetings

Regular meetings

Buddy system

PROJECT DELIVERABLES
T1 2021

Data Gathering

Data Cleaning

Data Visualisation

ML Algorithms

Predictions

Web Portal

PREVIOUS WORK

Web portal development

Geelong City Population Prediction

Insights from Crash data for Geelong

Visualizing insights from Real estate data and predicting the real estate price in Geelong

Crowd monitoring using Wi-Fi connected devices

Smart parking system data analytics

Key deliverables for T1 2021

Web portal unification and live deployment

Real Estate feature optimisation

Air Quality prediction from weather forecasts

Air Quality Team

Air Quality Feature - Goals

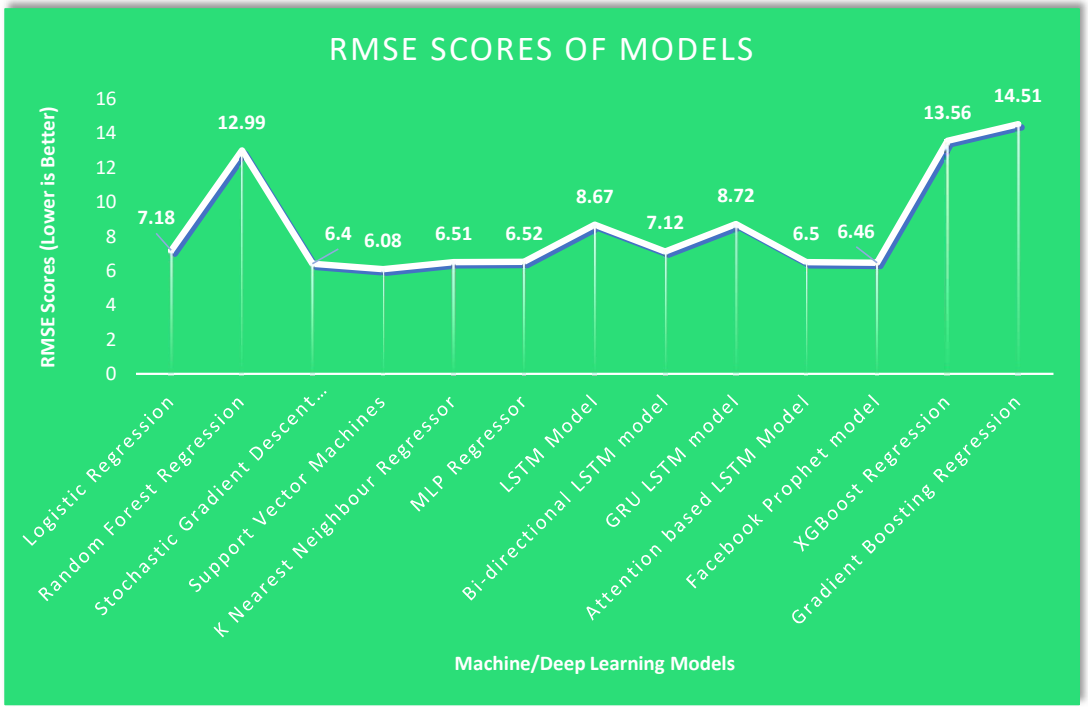
Data exploration and Feature selection

- Exploratory data analysis for weather and air quality indexing has been completed.
- Explored additional data from Barwon Health and Pollen grain count.
- Investigated impact of bush fire data on Geelong AQI
- Interesting relationships and correlations identified:
  - Main target variable of PM_{10} selected for prediction
  - Selected 4 predictor variables using feature selection (correlation).

Air Quality Feature – Modelling Technique

Machine Learning Models

- Built machine learning and deep learning models that forecasts the PM10 values.
  - Logistic Regression
  - Random Forest Regression
  - Stochastic Gradient Descent Regression
  - Support Vector Machines
  - K Nearest Neighbour Regressor
  - MLP Regressor
  - LSTM model
  - Bi-LSTM neural network model
  - GRU neural network model
  - Attention based LSTM neural network model
  - Facebook Prophet Model
  - XGBoost Regression
  - Gradient Boosting Regressor
- Chose Support Vector machines model due to low RMSE values



Air Quality Feature – Best Model

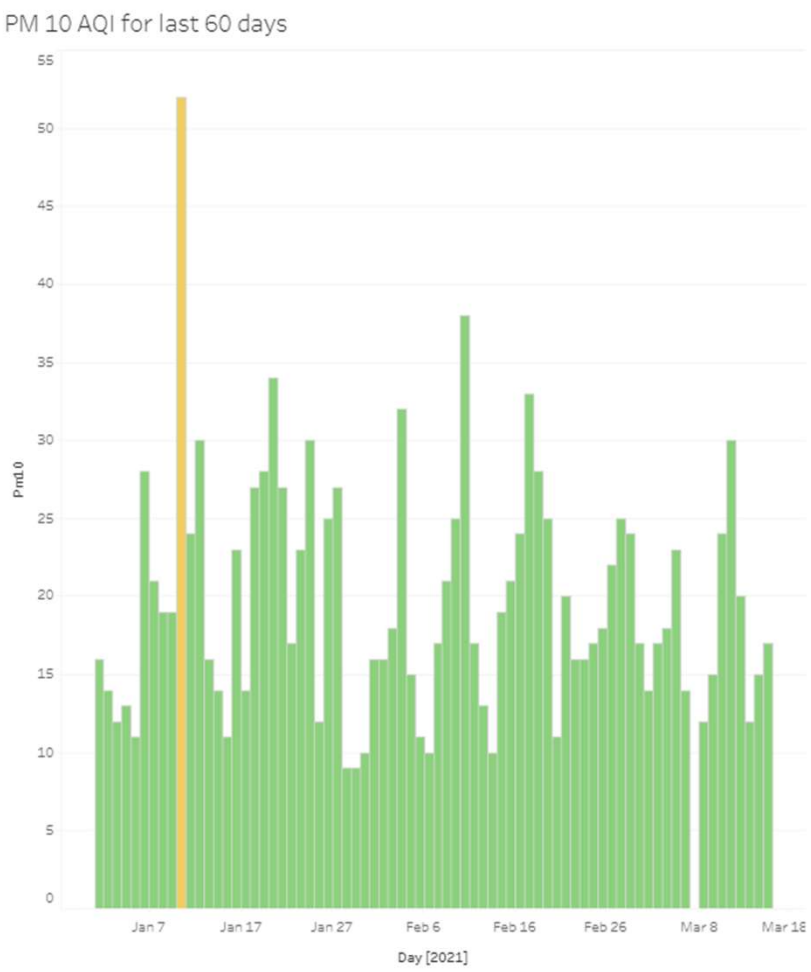
Best Model - Parameters

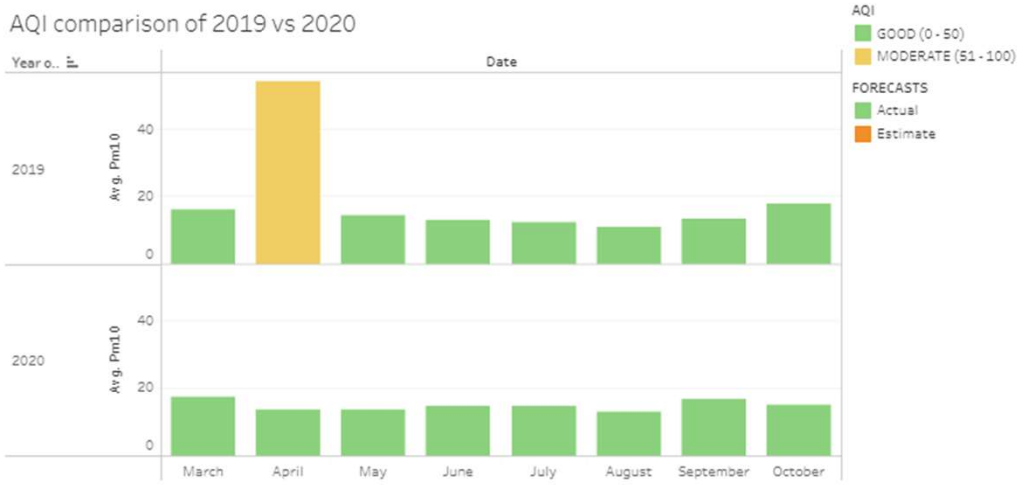
- Support Vector Machine – optimal model with less RMSE score of 6.08

Model	Input Features	Parameters
Support Vector Machines	• Maximum daily temperature • Rainfall • Speed of maximum wind gust • Direction of maximum wind gust	• Kernel : 'rbf', • C : 0.1 • Epsilon : 0.01 • Gamma : 0.1

Table : Best model parameters

Current Dashboard (AQ Feature)





Real Estate Team

Real Estate Feature Goals

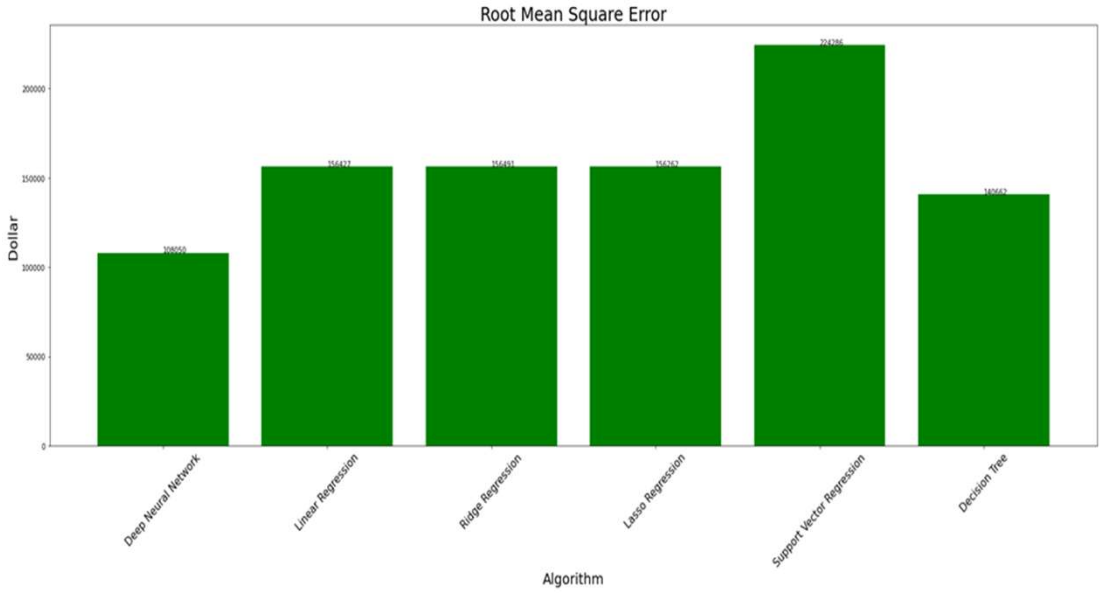
Goals For Trimester

- Optimize the existing Neural Network model
- Use other Machine Learning algorithms
- Visualize the impact of Covid-19 to the house price in Geelong

Real Estate Feature – Modelling Techniques

Machine Learning model

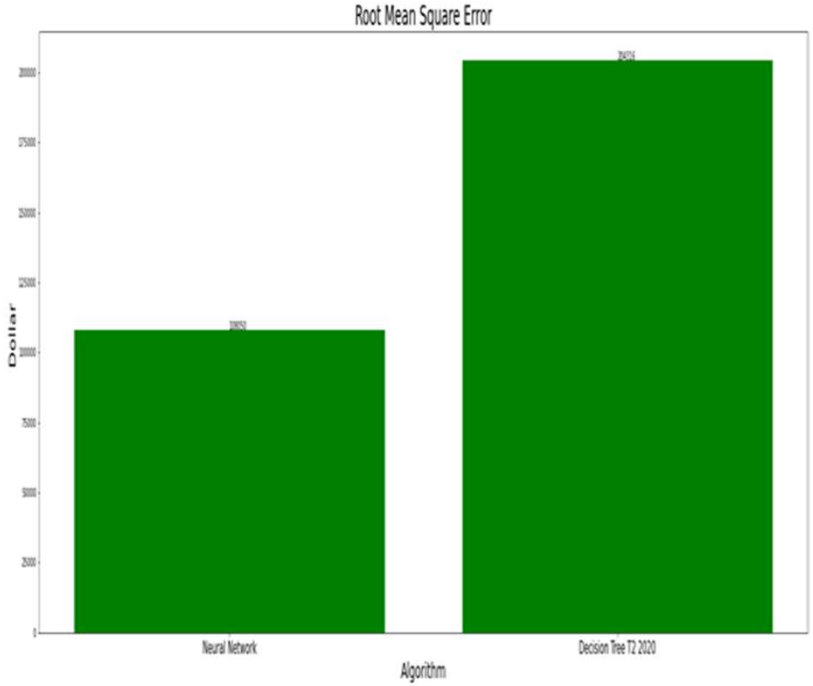
- Predict house prices in Geelong
- 6 algorithms:
  - Linear Regression
  - Ridge Regression
  - Lasso Regression
  - Support Vector Regression
  - Decision tree
  - Deep Neural Network (DNN)
- Root Mean Square Error (RMSE) is metric.
- Result:
  - DNN is the best model because it reaches the lowest RMSE value.



Real Estate Feature – Best Model

The optimal Neural Network model

- 13 inputs feature: Suburb, Land Size SQM, Bedrooms, Toilet, Car, School, Shop, Station, Park, Hospital, Rent Amount, Auction, Last Sell Price
- 3 hidden layers (128, 256, 128 nodes)
- Single node output
- Optimiser (Adam)
- Learning rate (0.01)
- Drop-out rate (0.5)
- Weight Decay regulizer
- Epochs (46)
- Results:
  - 50% reduction in RMSE compared to T2 2020 model



Real Estate Feature - Visualizations

Visualization

- Our visualizations include:
- The Map **and Ranking of Average Selling Prices of suburbs in Geelong.**
  - Visualize the impact of lockdown on house prices in Geelong.
  - **Median Facilities of in Geelong.**
  - Visualize the **Median House Price and Number of Covid-19 Cases in Geelong for the last 2 years**.
  - Median **House Price and Number of Lockdown days in Geelong**
  - Prediction of **the average house price in Geelong**.

Web Team Update

Web Development Goals

Goals identified for this trimester

- Unified Dashboard for features developed by D2I Geelong City
- Access to live website
- Optimize user interface for smoother interaction
- Tableau Report Integration for analytics
- User interface for ML features
- Infrastructure for prediction service

Web Development Goals - Achieved

Goals achieved in this trimester

- Migrated features Wi-Fi and Parking Data Analysis to single website
- Developed UI for real estate and AQI features
- Integrated updated tableau reports for analytics
- Integrated Google Places API to search nearby places for real-estate model
- Developed infrastructure for prediction service
- Deployed website and prediction application on Heroku
- Completed ground work for deployment on GCP

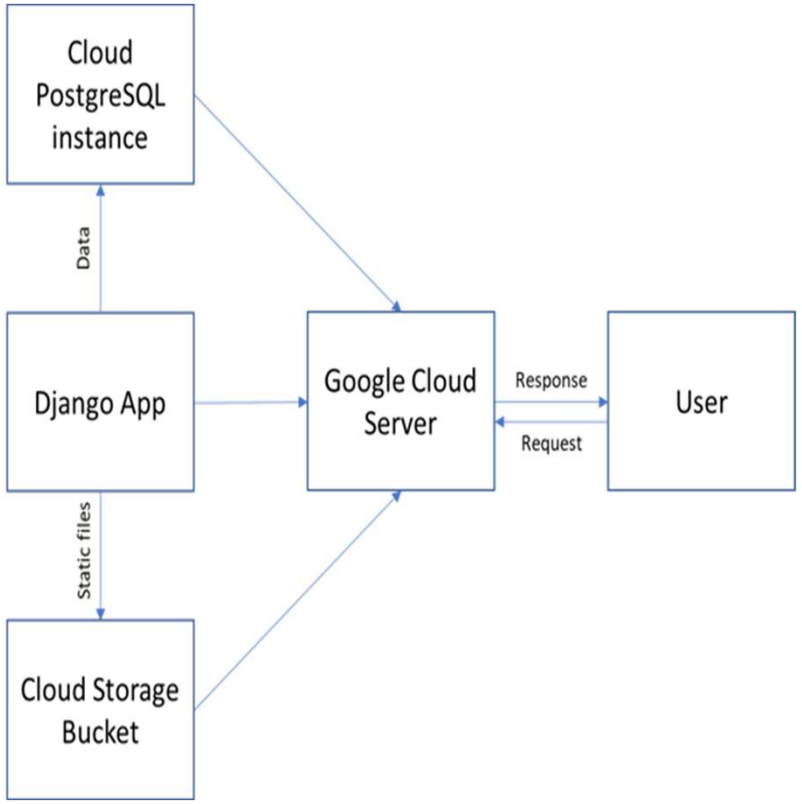
Machine Learning Model & Website Deployment – GCP (Researched)

Model Deployment Approach

- Deploy custom model on GCP AI Platform
  - Endpoint Creation for model deployment
  - Endpoint integration into Django application

Website Deployment Approach

- Create storage bucket for static files.
- Create PostgreSQL database instance for data storage.
- Modify Django app configuration settings to point to GCP



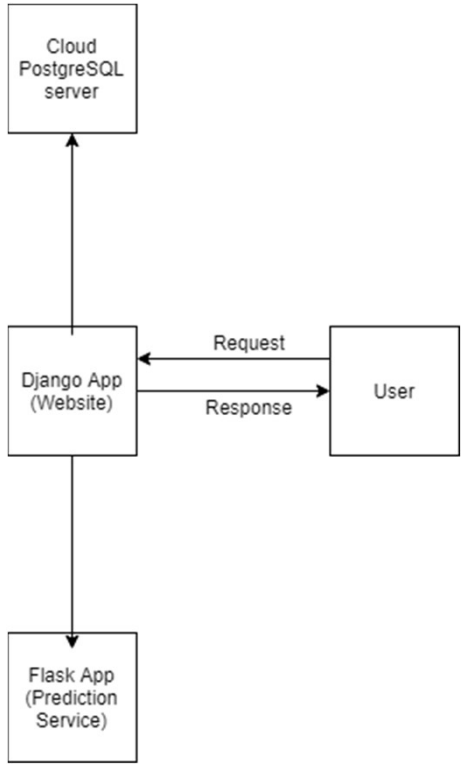
Machine Learning Model & Website Deployment – Heroku (Implemented)

Model Deployment Approach

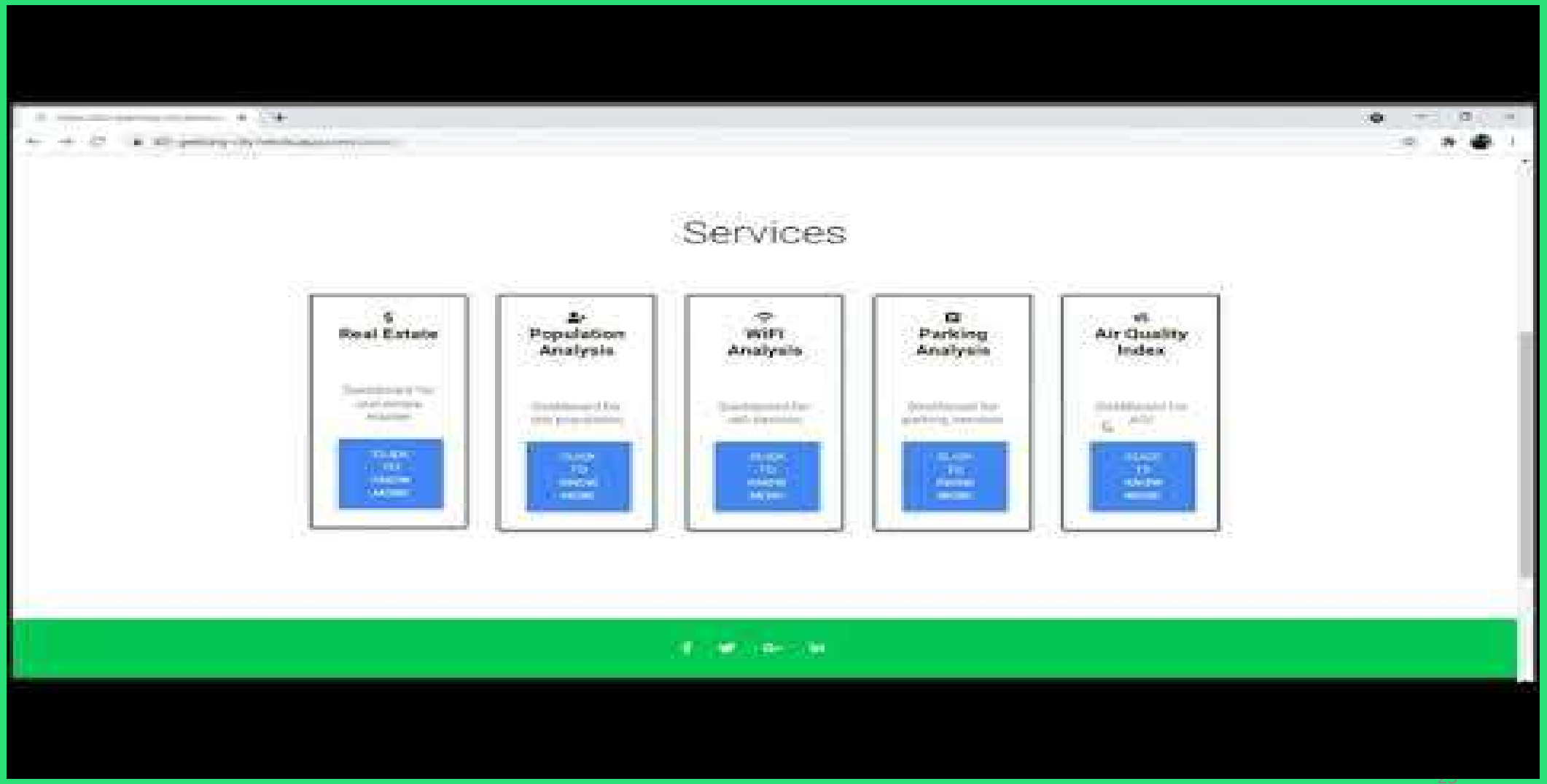
- Deployed custom model on Heroku Platform
- Created Endpoint for prediction from model using Flask

Website Deployment Approach

- Endpoint integrated into Django application for rendering results
- Deployed website on Heroku platform
- Integrated Heroku PostgreSQL database instance for data storage



Demo


Services
Real Estate
Population Analysis
WiFi Analysis
Parking Analysis
Air Quality Index

Plan For Next Trimester

Machine Learning Team (AQ and RE)

- For Air quality feature - Exploration of few more visualisations and integration of final tableau dashboard into Air quality prediction web page.
- For Real Estate feature - Add more data and add more features by google API & improve predictions.
- Build Deeper neural network and other models to reduce RMSE. Explore more visualizations on Tableau.
- Build data pipelines for ingestion and prediction
- Predict future price of housing unit based on the council investment like opening new shopping centres, parks in the suburbs

Plan For Next Trimester

Web Development Team

- Finalize deployment service as per instruction from University / DISC team
- Finalize text content for existing webpages to prepare website to be hosted in a live environment
- Build prediction services for the integration into client applications
- Minimize the page load time
- Provide beta access to client in order to receive and incorporate feedback

We love feedback!

Thank You!