Functional Specification - Car Deal Spotter

Name: Glen Devlin

Student No: 12524303

24/11/17

Table of Contents

Section 1 - Introduction

1.1 Overview

This project aims to build a model that can spot cars that are being sold at a good price. Using data scraped from carsireland.ie I will try to build a model using predictive algorithms that will attempt to find the value of a car and then use this to find cars that are being sold at a good price. Good priced meaning cars that are sold at less than they are valued. This project will involve building a scraper to get the data from carsireland.ie, cleaning and transforming the data then storing it in a database, building predictive algorithms to go over the data and finally creating an Android application to act as a front-end for the project.

1.2 Business Context

This project is being developed with data from carsireland as the terms of use for their site gives permission to use the data for non-commercial purposes and they do not forbid any sort of web scraping.

The app could be used by any users looking to buy a car. It will help them gauge the value of a car to stop them getting ripped off and help them save money.

It could also be used by anyone selling a car by letting them know how much the car they are selling is worth.

2. General Description

2.1 Product/System Functions

The project functions by taking data from the database that has been stored by the web scraper, then running predictive algorithms over this data to build a model, that can predict the price of a car. Any of the cars that are identified as a good deal by the model will be marked accordingly and passed on to the android app for a user to see.

2.2 User Characteristics and Objectives

The target market for the app is extremely varied. It includes anyone from around the age of eighteen to elderly people. Its main focus will be people who are looking to buy a car. The model will help these people by showing them how much they should be willing to pay and how much they should avoid paying.

Car buyers are a wide demographic and their technical expertise will vary. The front-end app however should be easy to use and even people with minimal technical expertise should be able to use it efficiently.

The aim is a simple android app that pulls the cars deemed good deals from the database and allows the user to refine their selection.

2.3 Operational Scenarios

The main operators of the app will be people buying cars and people selling cars. Buyers could use the app assist them in the purchase of cars. Sellers could use the app to assist them in valuing the vehicle they are selling.

Car buyer

Car buyers could use the app for assistance in buying a car. They would open the app and immediately shown a selection of cars. From this screen they

would be able to refine the cars they are being shown. They would be able to select the car's make, model, year etc. and the selection they are shown would adjust accordingly. What they are shown will be the information about the cars that has been pulled from the ad and a link to the original ad.

Car Seller

Car sellers would use the app to help them value the vehicle they are trying to sell. They could browse cars that are similar to the car they have for sale and this would allow them to estimate the value of the vehicle. There could also be an option to use the predictor model to value the user's car.

2.4 Constraints

Time

This project involves me learning a considerable amount over the next few months. I will have to learn how to scrape data and then how to clean and store the data. I will also have to learn how to implement predictive algorithms. Time management is extremely important.

Hardware

The only hardware I plan to use is my home laptop and an android device, most likely and android phone. I need to ensure I have all the required development tools installed and available. This includes Android Studio and some kind of database software. I may also use rent a server from Amazon to store my database.

Data

The data I take from carsireland.ie will need considerable attention. Often the data on these ads is missing various attributes. The data will need to be fixed in order for my algorithms to run correctly. I will also need to gather enough data so my algorithms have enough data to learn from.

Scraping

Scraping will be the first stage of this process and is crucial to retrieving the data. I will need to build a robust scraper and this may need to be redeveloped down the line if CarsIreland change the layout of their website.

Prediction Algorithms

I have very limited experience implementing machine learning algorithms and may need to learn a lot more before I can create a model that can retrieve good results.

Section 3 - Functional Requirements

3.1 Web Scraping

Description

To gather the data I will need to complete this project I will construct a web scraper. It will scrape carsireland.ie to retrieve the information for car ads. It will comb through the ads and take down the cars data and prepare it in a format so it can be stored in the database. The car information will be stored in the database along with an id that is contained in the URL. This will help me when performing multiple scrapes as it will mean I won't scrape the same car twice.

The web scraper will be built in python and will make use of the Beautiful Soup library. Beautiful Soup is a library for HTML parsing. The scraper will also perform some initial data cleaning and transformation.

Criticality

All other parts of the project depend on this section. Without the data I cannot move ahead to any of the other sections.

Technical Issues

If the CarsIreland website changes after the scraper has been completed, it could prove to be a real issue when performing future scrapes as the scraper might no longer function correctly.

Dependencies

The scraper will depend on the python library Beautiful Soup, a connection to the CarsIreland website and a database to store the data it collects.

3.2 Database

Description

Data gathered by the scraper will be stored in a database. The data will be cleaned after it has initially been put in the database. It will likely contain two tables. One for historic data the data analysis will use and one of cars that are currently available on CarsIreland. The database will use MySQL.

The database will be connected to all other sections of the project. It will take data from the scraper and provide data for the predictive analysis and the front-end app.

Criticality

How the data is kept is very important as better data will help me get be better results when I am at the stage of building the predictive model.

Technical Issues

The database will be stored on my home PC. There should be no technical issues.

Dependencies

The database will depend on the scraper to get data for it to store.

3.3 Predictive Algorithms

Description

Building the predictive model is the core part of this project. The model will analyse the data stored in the database and will then be able to predict the value of a car. It will most likely be based on linear regression/multiple linear regression; however other predictive methods will be tried to look for the best results.

The model will evaluate every car that comes into the database. If it decides that the car is being sold at a good price it will flag the car in the database as such.

The model will also be made in python. I will make use of the Pandas, Numpy and Scikit-Learn libraries. These libraries provide facilities for data structures and machine learning algorithms.

Criticality

This is the crucial part of the project. The model will need to be made well in order to successfully predict the value of a car.

Technical Issues

In order to predict the price of a given car I will need a large amount of data on cars that are similar. This could be a problem if there are not enough cars that are similar being sold.

Dependencies

It will depend on the data that is stored in the database and the Pandas, Numpy and Scikit-Learn libraries.

3.4 Front-end Application

Description

The front-end application will be a simple app for android devices. It will contain only the cars that are considered good deals and drop down menus to allow users to refine their selection.

Criticality

This is how end users will the application. It is important that it is intuitive to use and that it looks good.

Technical Issues

This app will be developed using android studio. In the past I have run into problems using the emulators that come with android studio. To get around this I will likely use an external emulator such as BlueStacks.

Dependencies

As this is the last stage to be developed it is dependent on all other areas of the project. Most importantly however it will need the database to pull car information from.

Section 4 - System Architecture
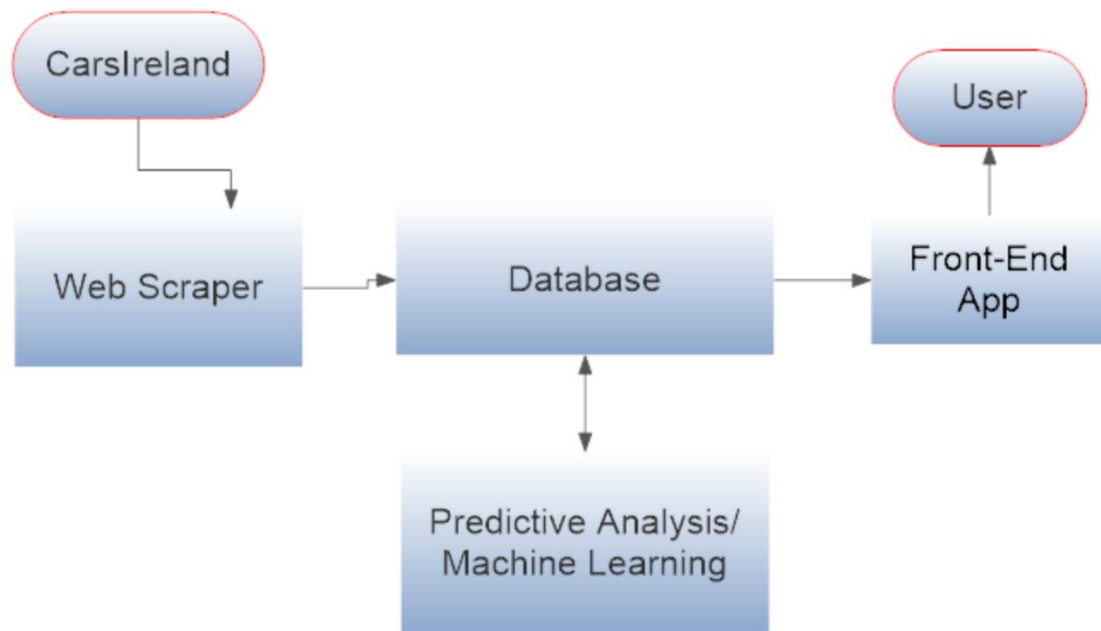
4.1 Web Scraper

The web scraper will gather data from the CarsIreland website and store it in the database. It will be made using python and will use the Beautiful Soup library.

4.2 Database

The database will be a MySQL database will take the data that it scraped from the web scraper. The data will be kept here so that it can provide information for the predictions and so that it can provide the actual ads to app.

4.3 Predictive Analysis/Machine Learning

To build a model that can predict the value of a car I will use python and the Scikit-learn, Numpy and Pandas library. It will only interact with the database,
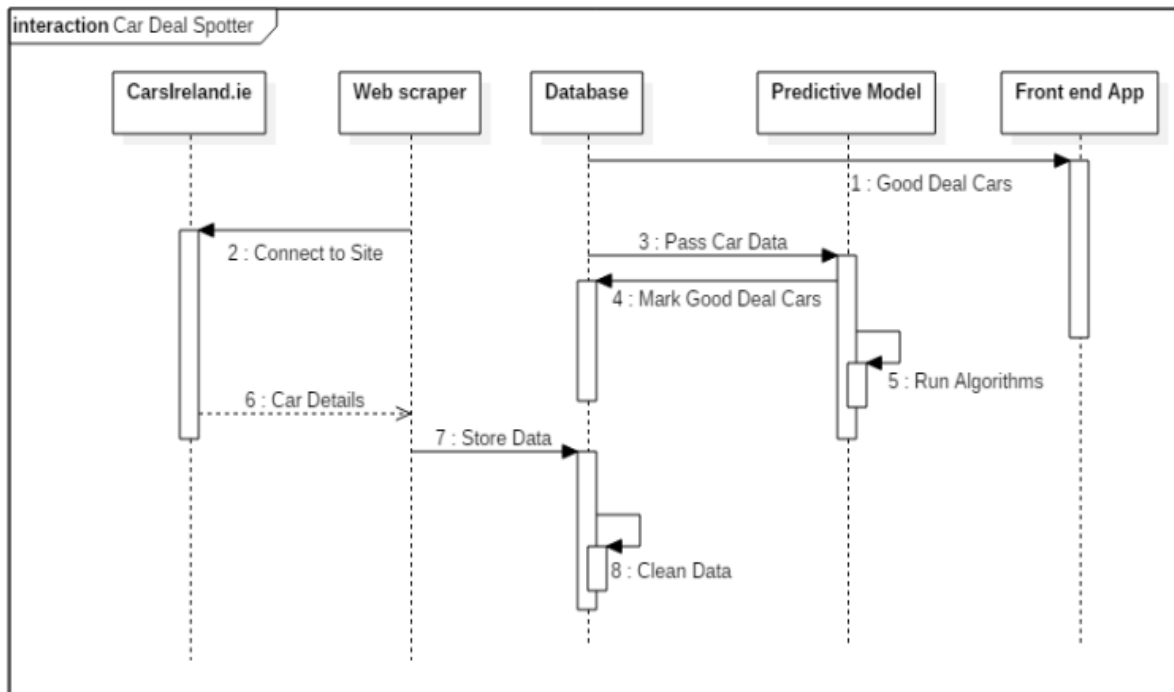
performing analysis and machine learning algorithms. It will only slightly alter data in the database.
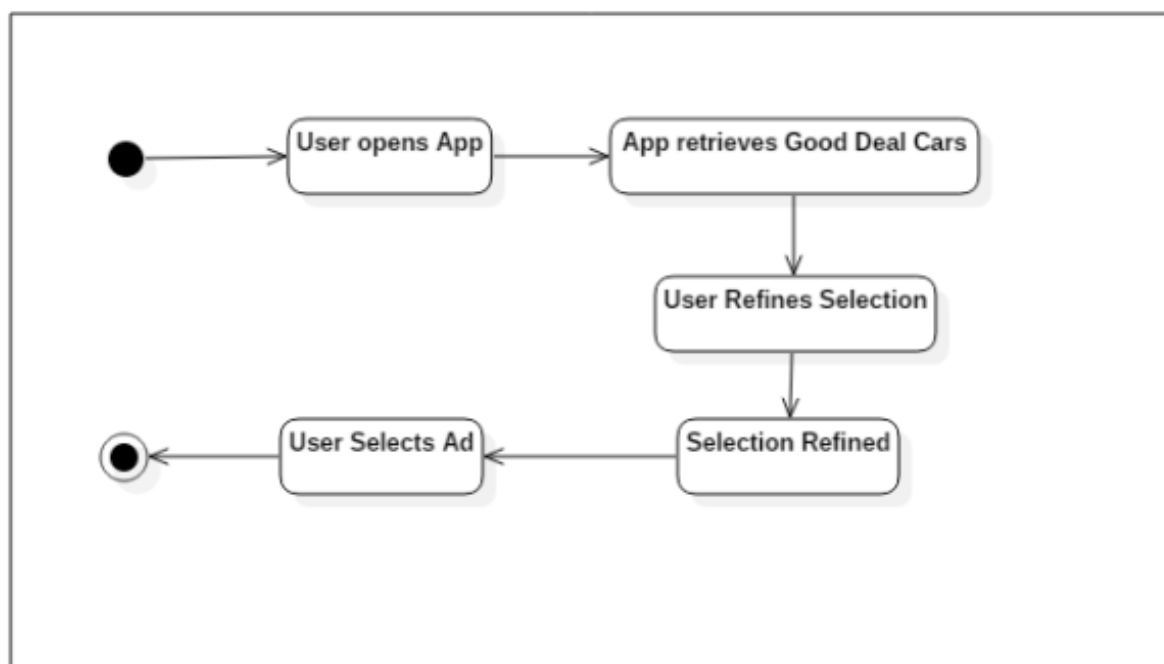

4.4 Front-End App

The front-end app will be an android app built in Java. It will pull information from the database to show the user.
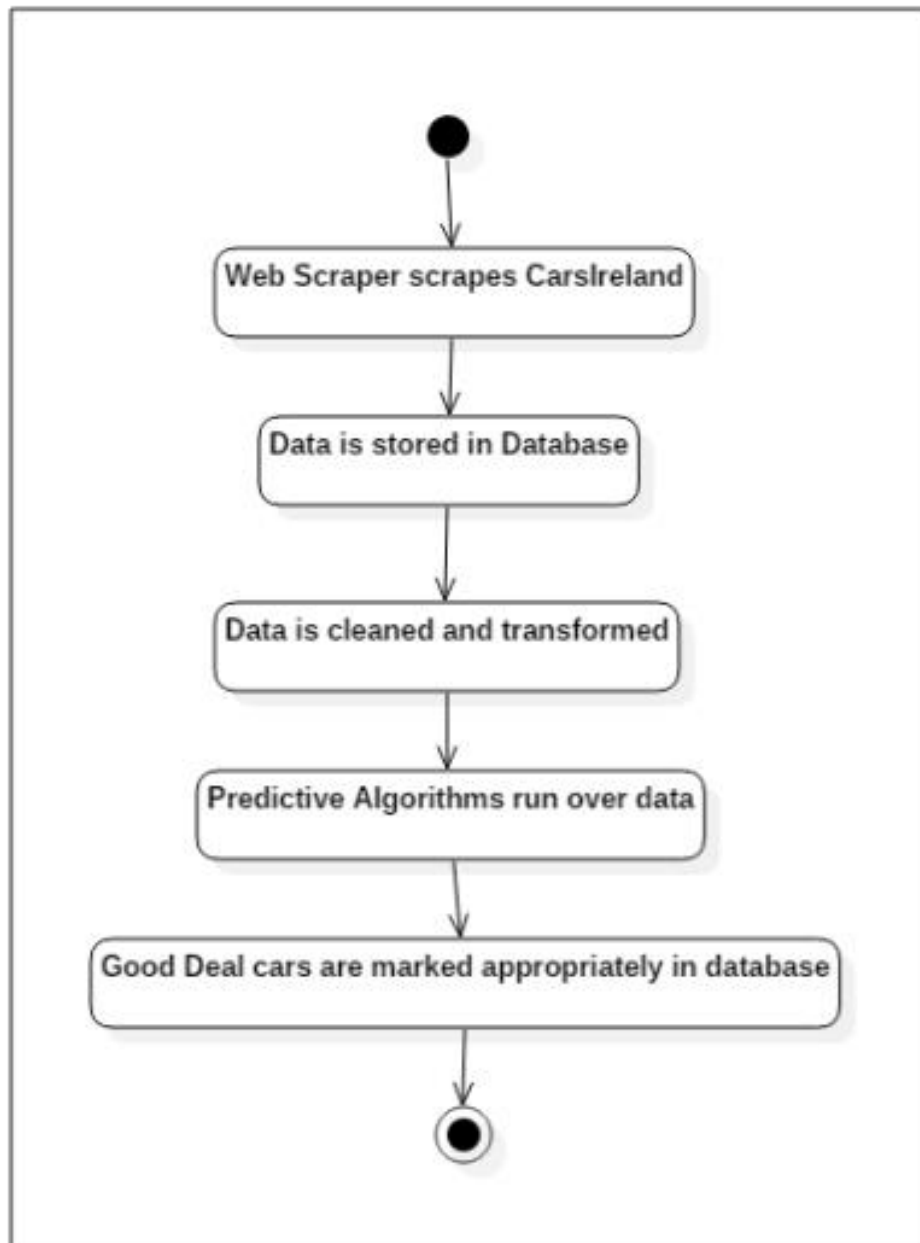
Section 5 - High-Level Design

Sequence Diagram

Activity Diagram – User

Activity Diagram – Web Scraping

Section 6 - Preliminary Schedule

6.1 Major Tasks

Web-Scraping

First I will have to build the web scraper. Next I will have to scrape the website; this could be a slow process as I may have to stagger the scraping. I may also have to perform multiple scrapes at later dates

Database

After the scraper is working it will need somewhere to store the data. Data may need extensive cleaning so I may have to write small scripts in python to clean the data.

Building Model

I will need to test out different machine learning algorithms to see which ones find me the best results. After I have found the method I will use I then begin to train it using all of the data from the database. I will then use the algorithm to find the cars being sold at a good price.

Front-end app

I will develop the app after all other stages have been completed.

Testing

Testing will take place though each stage of the project with additional user testing taking place after the front-end app is finished.

6.2 Schedule

Task Name	Start Date	End Date	Dec 10	Dec 17	Dec 24	Dec 31	Jan 7	Jan 14	Jan 21	Jan 28	Feb 4	Feb 11	Feb 18	Feb 25	Mar 4	Mar 11	Mar 18	Mar 25	Apr 1	Apr 8	Apr 15
Build Scraper	12/14/17	12/30/17	▓	▓																	
Retrieve Data	01/21/18	01/31/18							▓												
Setup Database	02/01/18	02/07/18								▓											
Fix and Store Data	02/08/18	02/14/18									▓										
Test Algorithms	02/15/18	02/28/18										▓									
Implement Model	03/01/18	03/24/18												▓							
Build Front-end App	03/25/18	04/07/18																▓			
Testing	04/08/18	04/21/18																		▓	

Section 7 - Appendices

External Libraries:

- BeautifulSoup
    - o https://www.crummy.com/software/BeautifulSoup/
- Scikit-Learn
    - o http://scikit-learn.org/stable/
- Numpy
    - o http://www.numpy.org/
- Pandas
    - o https://pandas.pydata.org/