<PiMyRide>: <David O'Regan >

Version 2.3
December 16th 2013

UI Contacts:
David O'Regan (david.oregan7@mail.dcu.ie)
10331017

1. Table of Contents

The table of contents is empty because you haven't selected paragraph styles to appear in it.

2. Introduction

- Purpose
- Definitions, acronyms, and abbreviations
- Business Context
- Overview

2.1 Purpose:

Overall CarDealz has three main goals, be able to gather information from a website(carzone.ie), trend the cars data compared against an average and finally relay that information to the end user. Our target audience would be car enthusiasts or users that want to buy a car but lack knowledge of the market. The product will be internally designed by David, all testing done will be in house until the end of the product development.

2.2 Definitions, acronyms & abbreviations:

App: This is the front end application the user will use to view "deals" as they are predicted by the web server. Examples would include an iPhone application/ Android Application.

Machine Learning: A branch of artificial intelligence, concerns the construction and study of systems that can learn from data.

Linear Regression: Is an approach to model the relationship between a scalar dependent variable y and one or more explanatory variables denoted X. This is how we will "predict" the best deals given a large amount of information.

Web Scraping: Is a computer software technique of extracting information from websites.

Web Server: The term **web server** can refer to either the hardware (the computer) or the software (the computer application) that helps to deliver web content that can be accessed through the Internet.

2.3 Business Context:

The business context for CarDealz is within two areas;
1) Single users
2) Carzone itself.

There are no business's sponsoring this project but I have contacted
   carzone.ie and if the project is successful they would be happy to
   implement it as a feature of their already standing application.

2.4 Overview:

The PiMyRide project consists of three main parts;

The first main part of this project is to develop a web scraper that will be able to
   successfully pull large amounts of data off carzone.ie. The main challenges of
   this part will be to implement it fairly so that while I get all the data I need, car
   zone is not bombarded with requests. Data to capture will be things like Car
   name/model/year/price/engine size/mileage etc. The main qualities that will
   affect the average sale price.

The second main part of this project is once the data has been scraped from the
   carzone.ie website, it needs to be stored into a web server. Once stored, the
   machine learning linear regression algorithm is applied to generate the best
   "average" price per car classification.

The third and last part of the project is to display that data(good deals) to a user
   via a front end application on an Android system.

3. Overall Description

- Product Functions
- PP(User characteristics)
- Operational Scenarios
- Product Constraints

3.1 Product Functions:

The general functionally of the CarDealz system is to spot* the best *deals as they appear on the website of our choice(carzone.ie). This is done by first "scraping" all the data we need from the website and storing it in an independent web server. Once the data we need is stored, a machine learning algorithm called linear regression will be used to generate the best average price for a car based on the information we have. Then once we have all this data and a general *best price for a given car category, we can watch the website and notify the user through an application when a deal is found. A deal will be a any given car that is below the best average price we have calculated for its class.

Example: The best average for a 2006 BMW M3 with leather seats and 60,000 miles in 4600 euro. If one is is seen for less than that price, the user will be prompted to view the possible deal.

3.2 User Characteristics:

Our user audience will mainly consist of motor enthusiasts, car zone itself and possibly people that know very little about the car market in general but want to shop for a great deal overall.

The main user community we will market to is a target demographic of people ages 21-30 primarily. The reasoning for this is 1) This bracket will have the most notable experience using Iphone/Andriod applications. We will not assume our users will have any programming knowledge but will anticipate for a reasonable grasp of using applications and/or actually owning a smart phone to run the application.

**The application will also be available as a web based application on a desktop but for marketability, a smartphone application is the winning choice.

Objectives & Requirements from the users perspective:
- User friendly application download and install.
- User friendly application interface
- Trending math's actually works i.e. Deals spotted are actually real possible deals and not bugs
- Web scraping will not interfere with application interface i.e. application does not crash during information download
- Search function(Search by price limit, year, car make etc)
- Push notifications for new deals that appear in real time
- Possibility to trend by preferences*
- Exclusive user log in's

3.3 Operational Scenarios:

The three operational scenarios will encompass a broad aspect of what a user will experience while using the CarDealz product:

1) Our user Carl(21 year old male) will download our application to his android phone, he follows the install process as directed and the application loads to his smartphone screen. After the download is finished, the application icon is highlighted and Carl is able to press it. Once Carl presses the application icon the interface begins to load and brings Carl to the start screen of the application.

2) Our user Nastaha(31 year old female) has already downloaded the CarDealz application and wants to use it to find the best deals on car zone as of now for the Nissan Micra brand in her price range. She will open the application, use the search function specifying the car model she wants to search and her price range. The application will search the database and return cars that are deemed "a good deal" for her price range or under.

3) Our user David who already has the application downloaded and has been using the application for some time(> two weeks) opens application to find a push application from web server to inform him of new possible deals avail be to view.

3.4 Product Constraints:
- Hardware Constraints
- Software Constraints
- Various Constraints
- Assumptions & Dependencies

3.4.2 Hardware Constraints:

The CarDealz system will be designed to run namely on Android based smart phones to begin with. A major hardware constraint is the need for a licence to development a IOS version of the application that be tested on a live server.

The system will then be constrained by the following hardware:
- User smartphone - Will run on Andriod(IOS is a possibility given ample development time)
- Web server memory capacity - More memory = greater data storage and larger/ faster results
- Processing power of the web scraper - The better the processor the we scraper is using the more/faster data can be collected.
- Wifi/network connection - Needed for application download, server communication and push notifications

3.4.3 Software Constraints:

The CarDealz system will be restrained through which software processes and applications are implemented to both retrieve and analysis the data.

The system will then be constrained by the following software:
- Web scraper designed - The web scraper being used needs to be built, thus how effective the scraping is will be dependant on the quality of the development.
- Trending algorithm - The accuracy of the data provided to the user will be dependant on how effective the trending algorithm(Liner regression) is designed.

3.4.5 Various Constraints:

- The system will depend on a the user having access to a smart phone.

- The system will be dependant on wifi/3g access to transmit and receive the data needed for it to run.
- The system will depend on how well the end user understands the use of an application i.e. download and use via their smartphone.

3.4.6 Assumptions & Dependencies:

Certain components of the CarDealz system contain assumptions and dependencies. Should any of these prove not to be the case during development, then this functional specisifaction will be review and edited accordingly.

Known assumptions are:
- Carzone.ie will not change their HTML design for the website as the web scraper relies on us understanding the static nature of the website.
- Carzone.ie are ok with us using their data and scraping at a reasonable pace. I have contacted the site and they are ok with the project as long as I keep the data requests reasonable. It is assumed this will not change.

4. Functional Requirements

- Web Scraper running
- Store data from scraper in web server
- Trend data stored using Machine learning
- Install frontend application
- Set up front end application
- Transmit data from server to application

4.1 Web Scraper Running:

- **Description:**

This is the first part of starting the system. The user will not have any involvement in this process, nor will they be aware of how it runs. The be scraper must be developed in house to gather data from the chosen website. The web scraper will run recursively, going from page to page on carzone.ie gathering the data we want. The scraper will be developed to target the most popular cars i.e. BMW, Nissian, Ford, Fiat.

- **Criticality:**

While simple, this stage is the crux of our entire system. The web scraper must run effectively to gather the data we need to manipulate for the user.

- **Technical Issues:**

The web scraper will be built in python and needs to be developed from scratch. Technically, the scraper needs to run recursively(going from page to page), mine the correct data(parameters) and finally not send too many data requests to carzone.ie over the scraping period(proper etiquette).

- **Dependencies:**

    Dependant on the developers knowledge, the structure of the carzone.ie webpages and finally that carzone.ie will allow us to scrape the data from their website.

4.2 Store data from scraper into Web server:

- **Description:**

Once the data is retrieved from the website carzone.ie from our web scraper, we will want to store the data into a online data base that can be accessed by our frontend application from any smartphone/PC. The data gathered from the website will be logged in either XML or CSV style files on the web server for further manipulation.

- **Criticality:**

This part of the system is again utterly critical. If the CarDealz system does not have an access to the stored data, we cannot trend it, or worse, display it to the end user in the form of "good deals".

- **Technical Issues:**

Technically, this web server will either be a dedicated machine or a bought VPS. Both will act and run fine for our needs, though a VPS is the more likely option.

- **Dependencies:**

Hardware needed to build web server, or a subscription needed for VPS.

4.3 Trend Stored data using Machine Learning:

- **Description:**

Once the data has been stored in our web server/VPS. We need to "trend" the data using machine learning(Linear regression) to establish two things, 1) What the baseline averages are for each category of car and 2) which cars appear to be good deals when compared against the averages.

- **Criticality**

This part of the system is perhaps the most critical of the entire operation. If the machine learning algorithm is not designed or implemented properly, the data shown to our end user will be wrong and not mean anything to them in relation to the applications original function i.e. spotting deals.

- **Technical Issues:**

To achieve this, the machine learning algorithm needs to be built to manipulate the stored data to spot the "deals" we want to display to the end user. This requires a knowledge of the machine learning concept and application, so that we can apply it properly to our data and generate real "deals".

- **Dependencies:**

Dependant on an accurate knowledge of machine learning, a effectively implemented algorithm and finally a large source of data to manipulate.
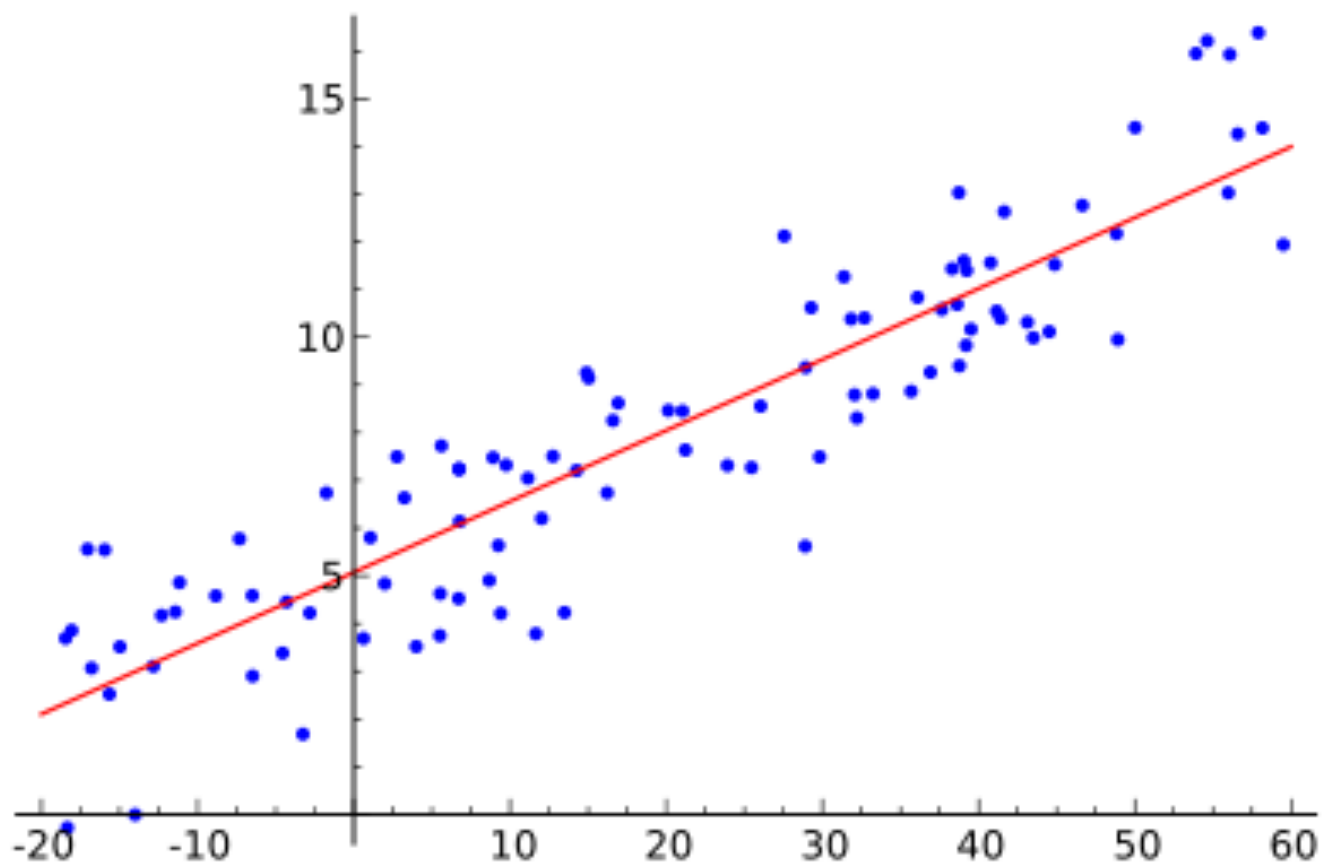
Linear Regression Model - Taking multiple values and generating a base line
  value.

4.4 Install frontend application:
  • **Description:**
  Our application will require a front end "display" portion to actually show the
    end user the data they want to see i.e. "deals spotted". The application will
    be developed for smart phone users, namely Android to begin with. A
    possibility for IOS porting if there is ample development time. The front
    end application will be downloaded from the android market place and
    installed.

  • **Criticality:**
    Again this stage is critical, our entire system depends on our data being
    displayed to the user in a readable, easily accessible fashion.

  • **Technical Issues:**
  Development relies on a understanding of Android application development.

  • **Dependencies:**
  Depends on user having access to smart phone, Wifi/3g for download and a
    understanding of how to use applications in general.

4.5 Configure Frontend Application:

- **Description:**

Once the application has been downloaded to the smartphone/PC in question,
 it will need to be configured. This may or may not include, creating a log in
 profile independent to each user, starting the software via some information
 input and possibly some data transfer from Web server/VPS to application.

- **Criticality:**

This part of the system is semi-critical to the developer as it refers more to the
 visual stimulus as opposed to the trends derived from the data gathered.
 Having a "independent" profile per user would be something will will strive
 for in development but it is not something that we deem absolutely needed.

- **Technical Issues:**

Technical issues will surround building a user independent log in so we can trend their own personal data. Depending on development time this may or may not be a feature.

- **Dependencies:**
  This stage is dependant on the amount of time we have to devote to developing a "independent" log in for each user and the possibility to trend data on a individual level.

4.6 Transmit data from server and display it via froent-end:

- **Description:**

This is the last stage of the CarDealz process, where the the data we have
trended will be displayed to our user via their front end application. This
will either be done automatically, via a selection search or we will possibly
allow the user more freedom with a blanket search bar.

- **Criticality:**

This part of the system is vital. It will be the entire bases for the machine
learning equations and allow us to actually display the dat we have
collected as the user wants it.
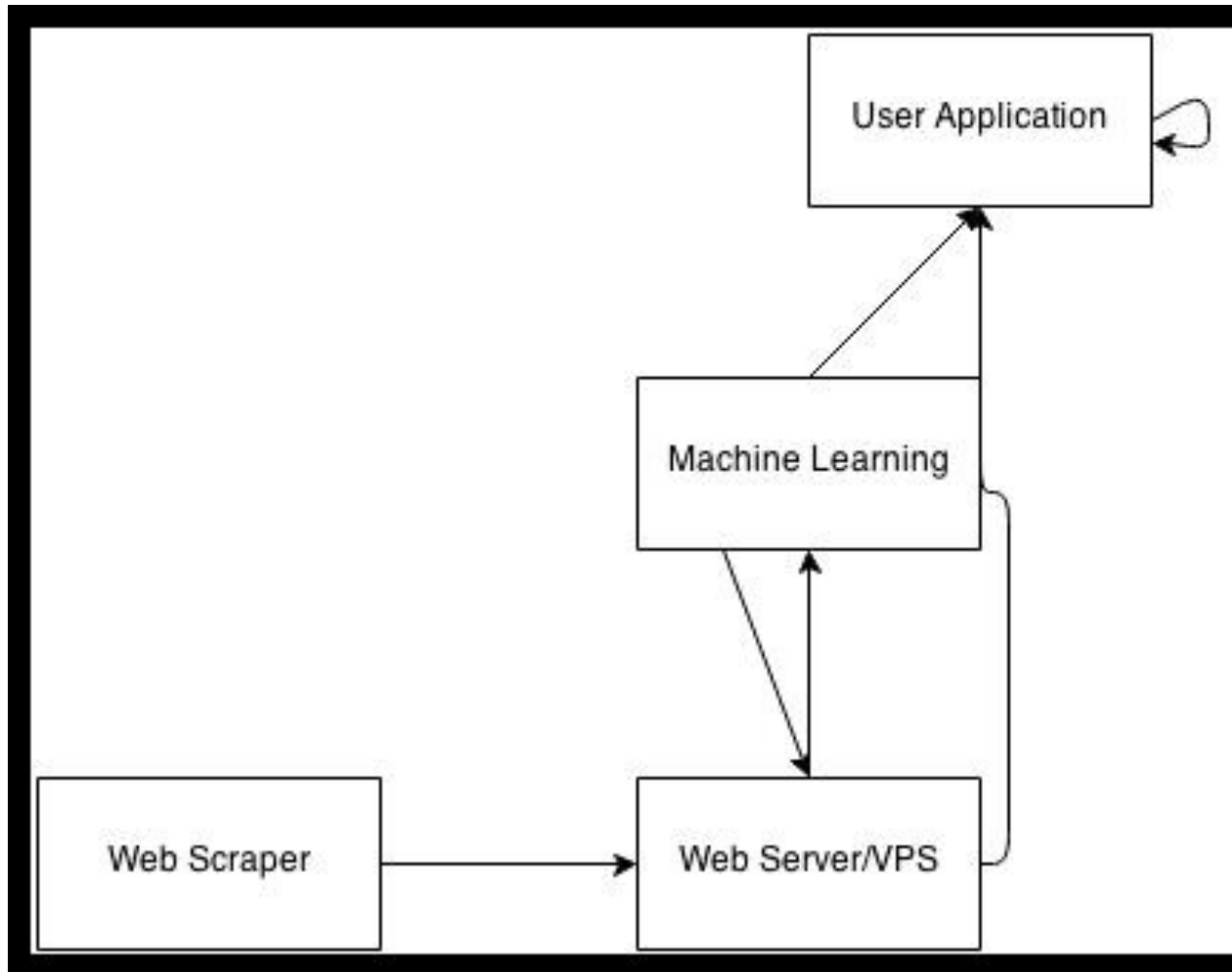

- **Technical Issues:**

The data will already have been trended by our machine learning algorithms
and waiting to be pulled down to our front end application. The most
important technical issue here is how we will display the data and how we
will let the user search the database. Depending on development time,
search options may be restricted in the beta and then expanded upon full
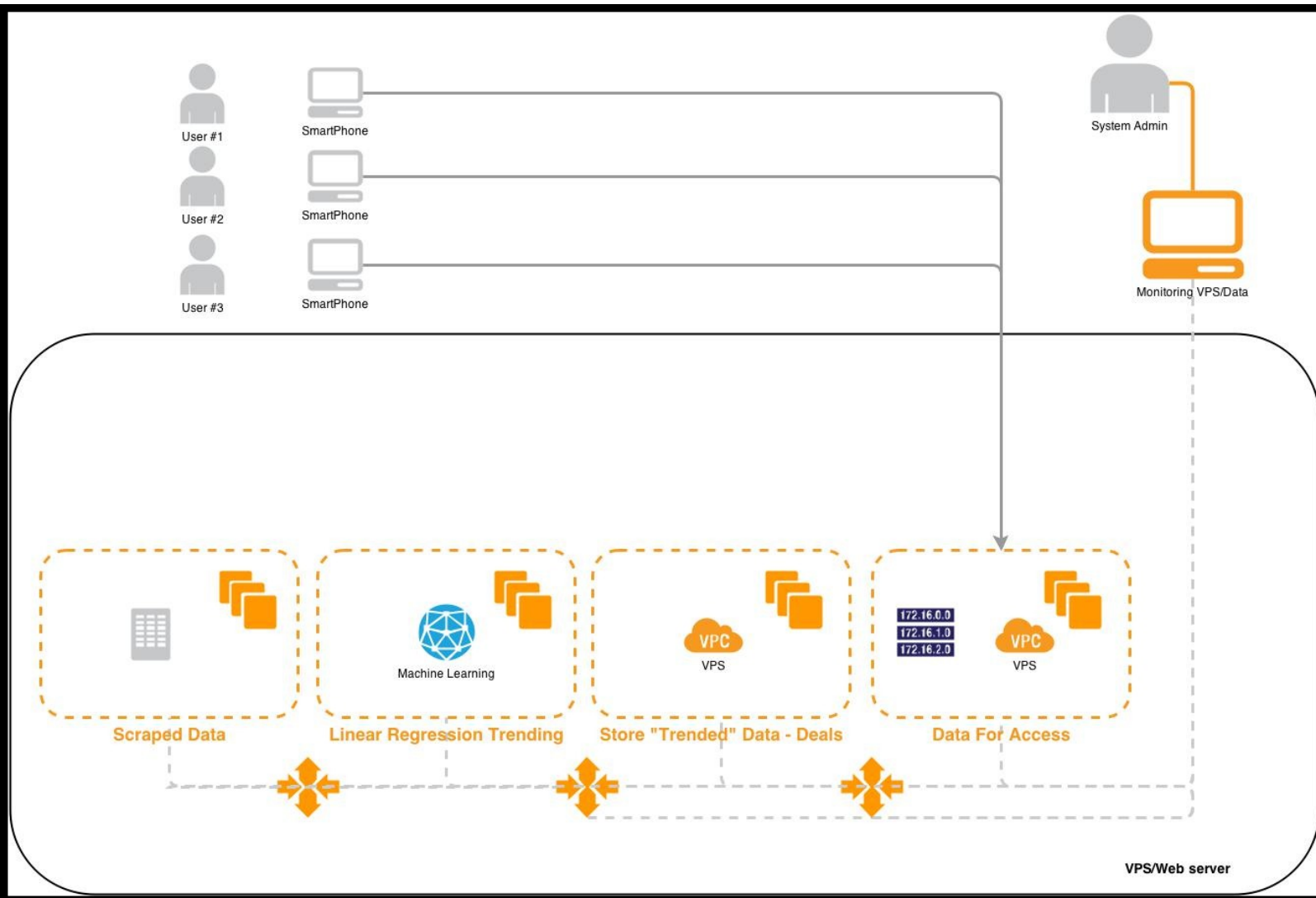development.

- **Dependencies:**

This will depend on the CarDealz being properly installed, the data being
accurate and the machine learning algorithms having working properly.
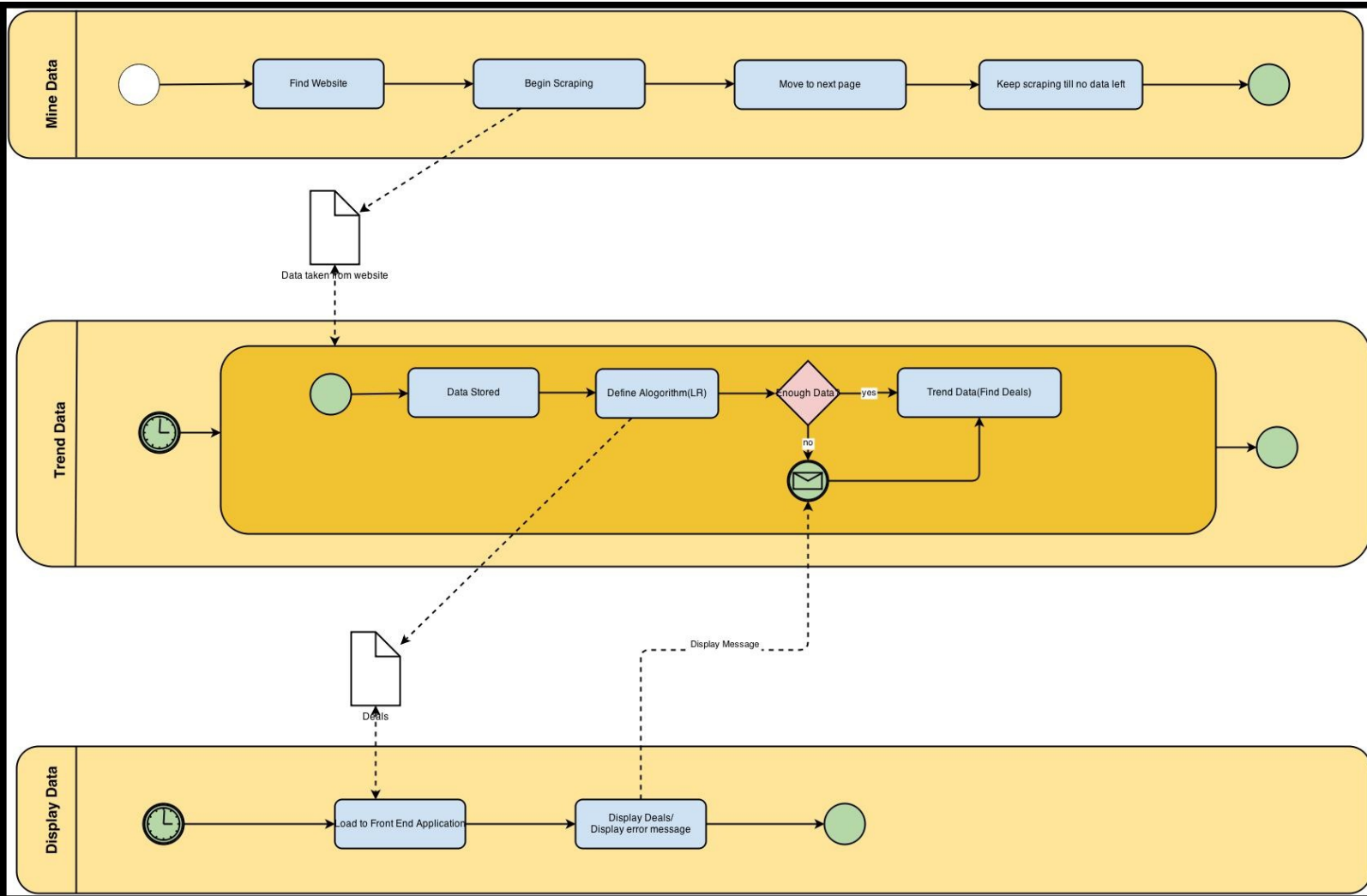
5. System Architecture

<System Architecture Diagram(Fig 2.1)>

<User Interaction Diagram(Fig 3.1)>

<Entity Relationship's Diagram(Fig 4.1)>

<Implementation Vs Abstract Layer Diagram(Fig 4.1)>
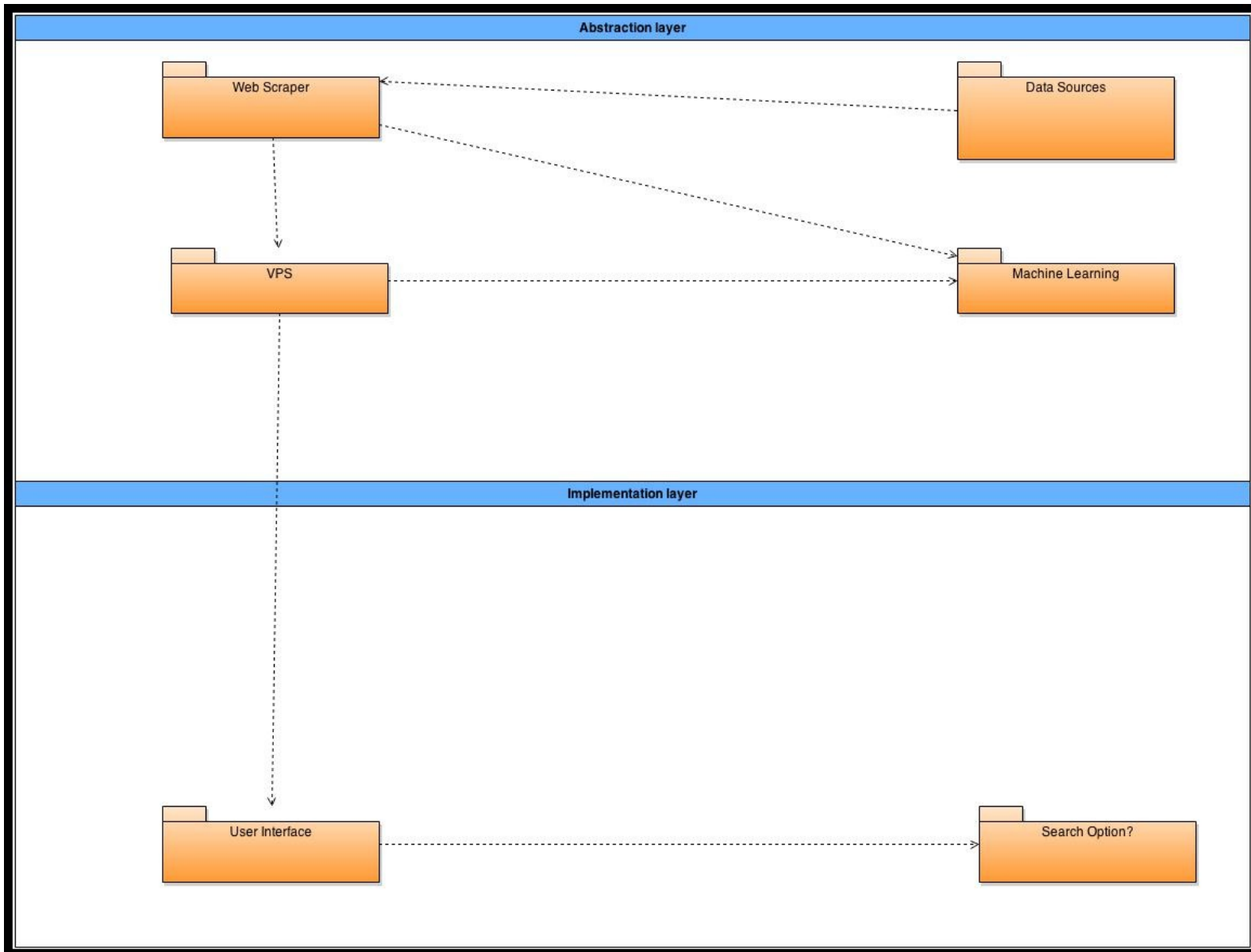
< System Architecture Description>

5.2.1 Web Scraper: This is the crux of our entire system. The web scraper will be a in house built application that will scrape* information from the website of our choice(carzone.ie). The scraper will be built in python and aim to scrape the website recursively but also fairly i.e. reasonable request rate.

5.2.2 Web Server/VPS: This the database we will use to store the the data we have scraped* from our website of choice(carzone.ie). It will either be a hardware hosted web server, or a purchased VPS(Virtual private server). The data collected will be stored here and then manipulated as needed by the rest of our system.

5.2.3 Machine Learning: This is our backend of the system and the core software part. The Machine Learning algorithm we will design(Linear Regression), will take the data we have stored in our data base(many items), and trend it by the parameters we provide to generate the average price for a given model of car, and then to spot* car's that appear to be below the average price for their category(Deals).

5.2.4 User Application: The final part of our system is the front end application, the actually application that will display the "deals" to our end users and possibly allow them to search the data base for specific deals. The application will be developed for Android to begin with, with a possibility for IOS porting towards the end of the system development.

6. High Level Design
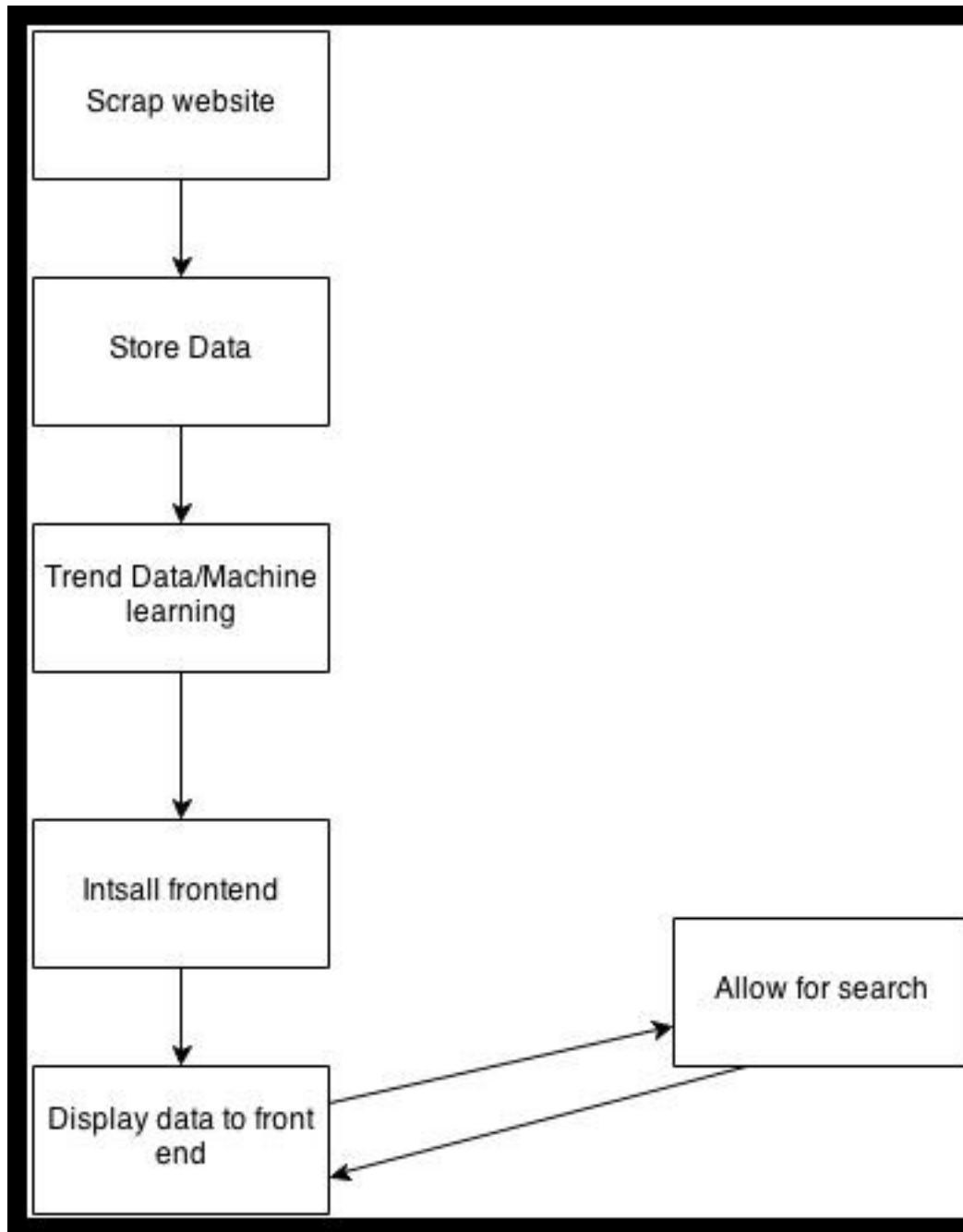
< High Level Description>

1.　　Scrap Website:

Our web scraper will gather the data we want to manipulate from the website of our choice.

2.　　Store Data:

Once the data has been scraped, the data is stored in a database hosed on either a web server or VPS.

3.　　Trend Data:

The data stored in the database will be fed into a machine learning algorithm(Linear Regression) to generate an average price for each car category, then any car's in said categories that are under the average price will be highlighted and stored in a separate list.

4.　　Install Front-End:

We need to install the front end on the end user's hardware platform of choice(smart/phone or pc).

5.　　Display Data to Front-End:

The data we have trended in our database now needs to be displayed to our end users application i.e. show them the deals that have been generated via the machine learning expression.

6.　　Search Option:
We will strive to give the end user a search option within the front end application to make finding relative deals to their interests more plausible. This will be dependant on time restraints.

7. Preliminary Schedule

Date	Main Objective	Secondary Objective	Misc Objective
16/12/2013	Functional Spec Submission	N/A	N/A
01/13/2014	Web Scraper Built	N/A	N/A
15/01/2014	Data Harvested From Website	Web Scraper tested fully	N/A
31/02/2014	VPS/Web server setup	Store Data in database	All data needed, taken from website
28/02/2014	Machine learning algorithm	Database stable	N/A
31/03/2014	Front end developed	Machine learning tested	N/A
30/04/2014	Data displaying on front end	Front end tested	N/A
10/05/2014	Search Bar addition	Possible IOS development	N/A
20/05/2014	Independent log in	N/A	N/A
25/05/2014	Main Testing	N/A	N/A
31/05/2014	Final testing	Independent Testing	N/A
3rd June 2014	Submission	All Testing Done	N/A
4th June 2014	Expo	N/A	N/A
5th - 10th June 2014	Demonstration	N/A	N/A

8. Appendix

Section	Date	Author/s	Issue	Pages
A	16/12/13	David O'Regan	Appendix A – Stategy	26 – 26
B	16/12/13	David O'Regan	Appendix B – Relivent orginisations & people	27 – 27

Appendix A - DESCRIPTION OF THE STRATEGY

Applications & Benefits:
- **Direct results:** *Users are provided instant feedback on what the best deals are for their chosen car categories.*
- **Public benefits:** *The common car owner is not forced to spent a large sum of money to discover the best deals within the second hand car community.*
- **Collateral benefits:** *None*
- **Scope of potential impact:** *Community and possibility for statewide.*

Critical elements:
- **Staff requirements: 1** *people*
- **Other requirements:** *Webserver/VPS, Smartphone*
- **Cost considerations:** *The program is costly to us as students but in the long run very cost effective*
- **Timing issues:** *Deadline allows for 0 weeks of full time work*

Potentional barries/obstecles:
- *Cost*
- *Development time*
- *Skills needed*
- *Hardware integration*

Appendix B – Relevent orginisations & people

Location	Primary Contact	Building Block
Dublin City University	*Daragh O'Brien*	*Project Supervisor*
Dublin City University	*Robert Hanrahan*	*Software development*
Dublin City University	*Joe Morris*	*Impartial feedback*
Beaumont Hospital	*Tom Byrne*	*Primary Tester*
Beaumont Hospital	*Fiona Croty*	*Secondary Tester*
N/A	*N/A*	*N/A*

*As the project continues this area will expand and be updated