**AI Based Cost Estimation System**

Project Report

|  |
| --- |
| by  CMPE 195A |
| Manmeet Gill  Sidarth Shahri  Manpreet Singh  Karthik Tella |
|  |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Dr. Vishnu Pendyala, Project Advisor  04/10/2019 |

# 

# **ABSTRACT**

# 

# **AI Based Cost Estimation System**

By Manmeet Gil, Sidarth Shahri, Manpreet Singh, Karthik Tella

Online shopping has grown massively over the past couple of years. For example, Amazon is one of the biggest websites for online shopping. Another very popular online market is buying/selling old and used items. Instead of having an old device lying around, it makes sense to want to get some money for it instead. However, there is no real way of knowing the prices of those items. It would not make sense to overpay for an older device. In the same sense, it would also not make sense to sell a valuable device for a very cheap price.

Manually visiting several seller websites and looking at various listings is not feasible if one wished to unload many products in a short time period. Looking for price quotes on the market for a variety of products is very time-consuming. Therefore, there is pretty much no effective way to find the average price of an item. Additionally, there are issues with sellers misquoting their prices, both accidentally and intentionally. Additionally, the price of used products on the market should degrade with time as the product deteriorates, occupies valuable virtual “shelf space”, and continues to be a burden on the seller. This follows that the seller would want to lower the price to get rid of the item faster or to make the sale.

Our product would essentially do most of the tedious work for the user, and return a value that is the average price of the given product based on the data gathered by our product. This would be possible through web scraping data from various used marketplaces such as Amazon, Craigslist, Ebay, etc. Then, using and filtering the dataset obtained from web scraping, AI models can be trained to predict price of certain item given some parameters based on average price of similar items in the dataset.

**Table of Contents**

# **Chapter 1 Project Overview**

# 1.1 Project Goals and Objectives………………………………………………….…….……….…3

# 1.2 Problem and Motivation……………………………………………………….…………..…...3

# 1.3 Project Application and Impact……………………………………………….…………..……4

# 1.4 Project Results and Deliverables……………………………………………….…………...….5

# **Chapter 2 Background and Related Work**

# 2.1 Background and Technologies………………………….………………….…………..……....6

# 2.2 State-of-the-art………………………………………….……………………………..……….6

**Chapter 3 System Design**

3.1 Architecture…………………………………………………………………………....….…...8

3.1.1 Model…………………………………………………………………………….……….….8

3.1.2 Controller………………………………………………………………………..…………...8

3.1.3 View…………………………….……………………………………………..……………..9

3.1.4 MVC Architecture Diagram…………………………………………………....……….……9

3.1.5 Backend Design Solution….…………………………………………………...……….…...10

3.2.1 Design Constraints and Challenges………………………………………………………….11

3.2.2 Design Solutions and Trade Offs…………………………………………....……………….11

**System Design**

# References………………………………………………………………………….……………...13

# **Chapter 1. Introduction**

**1.1** **Project Goals and Objectives**

The project’s main goal is to empower customers who are buying and selling items on online marketplaces for used goods such as Craigslist, Facebook marketplace and eBay. Our product aims to utilize advertisements from online marketplace as precepts to train a learning agent which utilizes the knowledge gained from these precepts to provide estimates and analytics about the availability and price of desired products.

**1.2** **Problem and Motivation**

The main problems with our project are mostly technical and legal. Technical problems include making sense of online advertisements which are mostly unstructured, web scraping from dynamic websites and hosting our learning agent. Legal problems mainly come down to the accessing rights on websites. Some websites are against third parties accessing their source code and benefiting from it, thus claiming them to be property. Such websites can be scraped but usually should not be scraped without the consent of the domain owner. Therefore, gaining rights or consent to access some marketplaces could be a problem. The main motivation of the project is to empower online customers and sellers given the vast and growing destinations online.

The needs for our project mainly include ad data, hosting space and source code/data access privileges to preexisting marketplaces. With online marketplaces for used products gaining popularity, the sheer number of advertisements could overwhelm buyers and sellers in making a decision. Our project uses web scraping and machine learning to analyze a big portion of these ad space to analyze common trends for a given product simplifying the process of researching and decision making for users.

**1.3 Project Application and Impact**

The goal of our project is to be able to return to the user an accurate price of an item on the used marketplace. Upon completion of our project, it will directly affect the used marketplace. Our project will also have an impact on academics, society, and industry. Firstly, in school students tend to buy used textbooks because they tend to be a lot cheaper than buying brand new. Thus, saving students hundreds of dollars. Our project would help with this process because students will be able to get a general price for the textbook, and then be able to find themselves the best deal. Students would also be able to find expensive materials such as: laptops, textbooks, and software for a good price with our project. This also applies for society. Technology is becoming more and more advanced, and with it prices have also been gradually going up. For example, the iPhone X costs around $1000 brand new. This is where the used marketplace comes in handy. Consumers can buy slightly used technologies for a fraction of the price, making some of these expensive technologies obtainable for the lower/middle class. This will also obviously have an impact on industry as used marketplaces take customers away from the new marketplaces. Industry will not take a direct hit in terms of revenue because in order for an item to be on the used marketplace, someone had to have bought it first. In the future if the used marketplace becomes very popular, it could take a large amount of customers away from buying new items.

**1.4 Project Results and Deliverables**

Upon completion of our project we hope to be able to present a clean front-end website/application, where the user is able to search for an item and receive an accurate price for that item. We will also be designing the back-end. There will be two parts to the back-end: data collection and data parsing. We will be obtaining data from various used marketplaces, and then parsing through that data to be able to calculate an accurate average price of a given item.

# 

# **Chapter 2. Background and Related Work**

**2.1** **Background and Technologies**

The project is divided into 3 main segments based which include data procurement, data analysis (learning) and interface. Data procurement mainly deals with accessing advertisements from online marketplaces. To make the process seamless our project will implement web scraping which is going to be done using Python 3 with the beautifulsoup and requests libraries. As a part of procurement, we will also structure this data and save it as (.csv) comma separated values format. In order to excel, we will be using Python 3 and Microsoft excel with the Pandas library.

Data analysis mainly encompasses our intelligent agent which takes the data from the previous step and train a model which will be written using Python 3 and the Tensorflow API. It will be a regression model. Using training data, we will fit the regression model. Price will be our Y variable while other variables being x’s.

Interface of the project mainly deals with our GUI and hosting. The GUI will be developed using JavaScript with ReactJs as a choice of library. Some of the other dependencies of GUI include npm, webpack and other third-party styling libraries. Hosting and backend will mainly be done on AWS with technologies such as dynamoDb and Node Js being used.

**2.2** **State-of-the-art**

According to a Deloitte Global research paper, Used cell phone market is one of the most growing markets which has an economic value of about 17 billion with an annual increase of about 80 million every year (Deloitte, 2017, p. 1). Our application provides estimate on price of the cell phone given the information about physical and internal attributes of the cell phone using machine learning algorithms on data from other advertisements on public marketplaces. Any product providing similar service for same domain i.e. cell phones is not yet available in the market. Dr. Zafar Khan and Dr. Asim, from University of Engineering and Technology, Lahore, has written a research paper which revolves around using classification to predict prices of new cell phones using dataset available on GSMarena.com (Asim, 2018, p. 2). Our solution will solve of problem of underpaying or overpaying for a used product.

# 

# **Chapter 3. System Design**

**3.1 Architecture Design**

We decided to use MVC architecture for our application. MVC is one of the popular software architecture design patterns. It will help us to maintain our application more efficiently. MVC architecture is consists of three main parts- model , view and controller. We will use ReactJS as view, node js and express js as the controller and mongodb(noSQL) as the controller,to build our application.

**3.1.1 Model**

We will use an open source nosql database MongoDB to build our application. Interaction with the MongoDB database is the model of our application. The model of our application will not be directly connected with the frontend part. When a user logs into their account on our AI Based Cost Estimation System, the controller receives the request. The controller, in our case, node js, will use the model to retrieve data from MongoDB and it will show the data in the view , the UI of our web application.

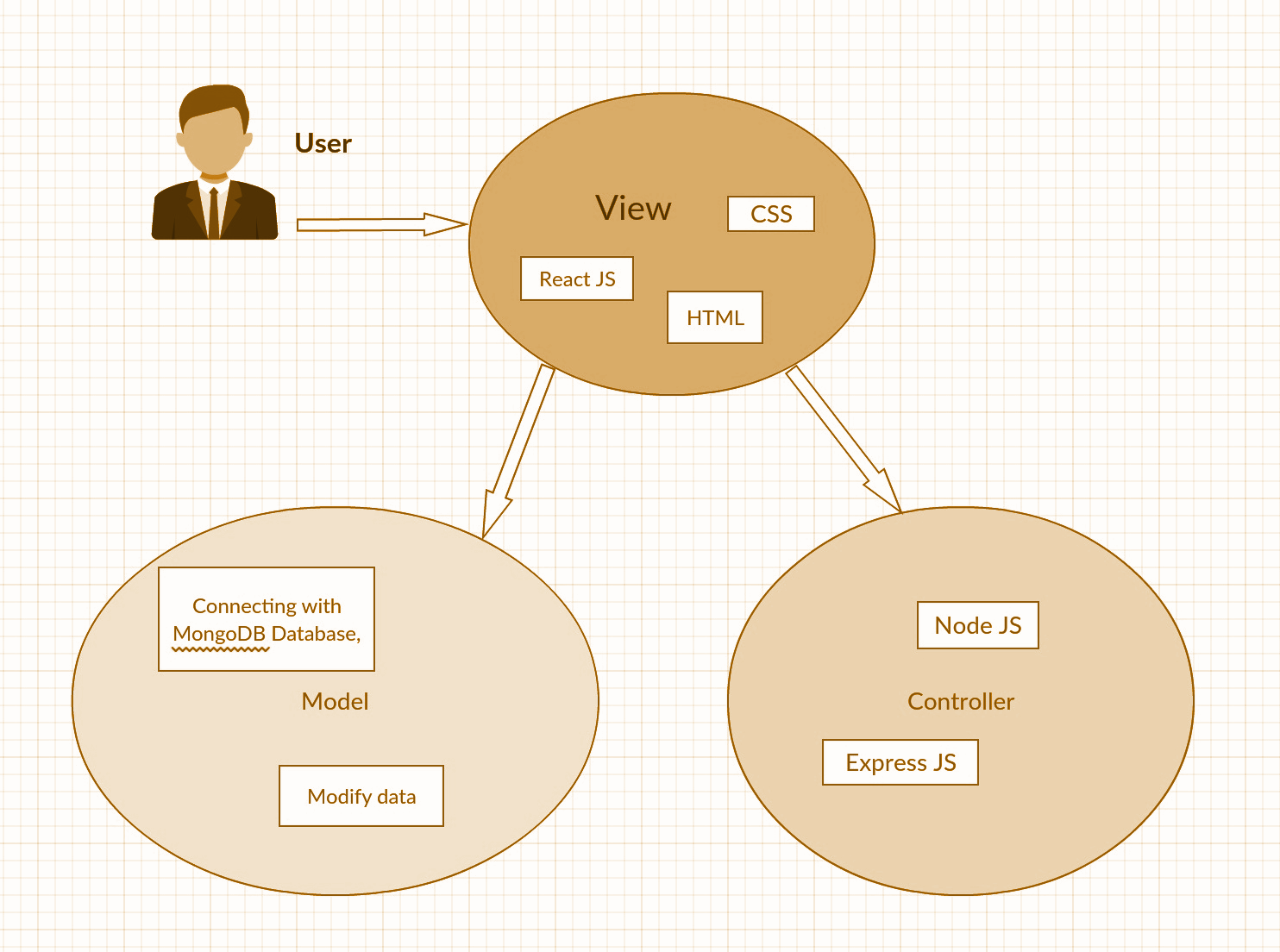
**3.1.2 Controller**

Node js is the controller in our application. We will use node js to process requests from the model to the view the data by getting parameters from view for the request. We will also use node js to get data from the model.

**3.1.3 View**

We are using ReactJS to build the front-end part. We will also use HTML, CSS and Bootstrap to create the UI of the application. The front-end part of our application is the view where users can enter parameters like condition, brand, memory etc. to get the cost estimate. The view sends this request to the controller.

**3.1.4 MVC Architecture Diagram**

****

**Figure 1:** Diagram shows how the user interacts with our MVC model.

**3.1.5 Backend Solution Design**

After using Beautiful Soup’s library to collect data from public marketplaces, we will store that data into Google Drive. Using the unfiltered data, we will filter out our data into labels like brand, model, memory etc. and store that into a MongoDB database which in JSON format which can easily be converted into Comma Separated Values (CSV) format. We will also preprocess our data by removing rows with empty values and data with extreme values like $0 or $1000. Next, we will split data into train and test datasets by 70% and 30% respectively. We will use algorithms like linear regression and k Nearest Neighbors (kNN) from libraries like Scikit Learn or TensorFlow to fit with training data. Then we will be using test data to test the accuracy of predicted values against the actual values. We will be using other machine learning techniques like feature engineering to improve our scoring metrics like accuracy score and F1 score. After we have attained a certain accuracy score, we will use whole original non-split dataset to train our model which will be accessed by node.js. Anytime a request will be made by node.js, our model will predict the cost (y variable).

**3.2 Design Constraints, Problems, Trade-offs, and Solutions**

**3.2.1 Design Constraints and Challenges**

The designs constraints and challenges for the project can be categorized based on the following perspectives:

* Economic: Being a student funded project the funds are limited to host data and websites.
* Resources: Since the project is heavily dependent on data, resource related challenges mainly come down to data procurement and data storage.
* Society and Environment: Getting a consumer base to adapt to the product resulting from the project would be challenging.
* Hardware/Software: The project mainly deals with two major languages Python and JavaScript. Python being used for data procurement, structuring and machine learning. While JavaScript being used for web hosting, interfaces, etc. Being able to create a communication channel between these separate entities would be challenging. Some solutions include the use of .JSONs or multipurpose API’s.
* Mathematical/Scientific: To reap benefits from a machine learning model at least 10,000 data points are needed for training purposes. Finding such data and structuring it would be challenging.

**3.2.2 Design Solutions and Trade-offs**

Our solutions for the constraints and challenges listed above:

* Economic: Since we are a student funded project and access to funds is low, we have to use a lot of free/cheaper alternatives rather than expensive products. For example, for our storage we will use the free storage given to us by either google drive or dropbox. Since the product is free, there will be a cap to the storage overall and a potential bottleneck on the access of materials.
* Resources: Data storage and data collection are our biggest challenges. Data storage relates to the solution discussed in the economic section with using free data storage methods. Data collection will essentially be run on our local machines for the time being, but eventually we will have to move to a server. Thus, the main trade off here is just the amount of websites and posts we will be able to scrape in a given time period.
* Society and Environment: The main challenge here is getting consumers to use the application. The easiest way to address this challenge is to promote the product to as many people as possible to gain some popularity and traction. Eventually the user base will develop, as long as the application does as intended.
* Hardware/Software: The main challenge here is using two languages and having them work together properly. This will require a lot of research, trial and error, and help from our advisor. We will have to ensure that our product is developed in a way such that it is maintainable and scalable.
* Mathematical/Scientific: Since 10,000 data points are needed for training purposes, this challenge ties back into our other challenges. This heavily depends on how well our program is developed, and the services we use for storage purposes as stated in the economic and resources sections.

**References**

Asim, Muhammad & Khan, Zafar. (2018). Mobile Price Class prediction using Machine Learning Techniques. *International Journal of Computer Applications*. 179. 6-11. 10.5120/ijca2018916555.

Deloitte. (2017). Used Smartphones: $17 Billion market you may have never heard of. *Deloitte Network Global,*(4), 3. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Technology-Media-Telecommunications/gx-tmt-prediction-used-smartphones.pdf.