RETAIL PRICE PREDICTION USING MACHINE LEARNING

### A Project Progress Report -II

*Submitted in the partial fulfillment for the award of the degree of*

# BACHELOR OF ENGINEERING

**IN**

**ARTIFICIAL INTELLIGENCE AND**

**MACHINE LEARNING**

### Submitted by:

### TANYA GUPTA

### (20BCS6600)

### TANUSH SHARMA

### (20BCS6596)

### SRI KUMAR DAS

### (20BCS6629)

### VEDANT

### (20BCS6624)

### Under the Supervision of:

### MONIKA SINGH



**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,**

**PUNJAB**

**FEBRUARY 2022**

**DECLARATION**

We, **‘Tanya Gupta’, ‘Tanush Sharma’, ‘Sri Kumar Das’** & **‘Vedant’** student of **‘Bachelor of Engineering in Branch Name’**, **session: \_2020-2024,** Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled **‘Retail Price Prediction’** is the outcome of our bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

**Tanya Gupta**

**Tanush Sharma**

**Sri Kumar Das**

**Vedant**

**Student UID’s:**

**(20BCS6600)**

**(20BCS6596)**

**(20BCS6629)**

**(20BCS6624)**

**Date: 19/04/2022**

**Place: Chandigarh university**

**ACKNOWLEDGEMENT**

We would like to earnestly acknowledge the sincere efforts and valuable time given by Professor Monika Singh. Their valuable guidance and feedback have helped us in completing this project. Also, we would like to thank our group members who supported each other in making and completing this project successively. Our sincere efforts finally produced something that we imagined. We would also like to thank faculty who provided us this golden chance. Last but not least we would like to thank everyone who supported and motivated us for pushing ourself beyond our strength.

Thank You.

**ABSTRACT**

Predicting retail prices of things is something that both businesses and customers are interested in. The price of a product at a certain target outlet is often related to the pricing of the same or similar products at neighboring competing outlets in a competitive setting.

Using four-vector autoregression models that incorporate the historical retail prices of the product at a target outlet and at rival outlets, this study predicts the start of day and current pricing of a certain product at every outlet in a given city. The models also contain the product's anticipated wholesale price. Three methods are discussed for identifying local rivals. A wholesale supplier is one that has price patterns that are similar to those of a target outlet. A basic autoregression technique outperforms the proposed models.

# Table of Contents

|  |  |
| --- | --- |
| Title Page  Declaration of the Student Abstract Acknowledgement  List of Figures  List of Tables (optional) Timeline / Gantt Chart | i |
| ii |
| iii |
| iv |
| v |
| vi |
| vii |

|  |  |  |
| --- | --- | --- |
| **S.no.** | **Topic** | **Page number** |
| 1. | Introduction | 1 |
| 2. | Literature Review Summary | 1-2 |
| 3. | Problem Formulation | 3 |
| 4. | Research objectives | 4 |
| 5. | Methodology | 5-7 |
|  | | |
| 6. | List of references | 8 |

## List of Tables

#### Table Title page

Literature Review Summary

## List of Figures

#### Figure Title page

Flow Chart

# INTRODUCTION

**1.1**

When it comes to appropriately pricing their items, the biggest challenge for retail teams is answering the question: What is a fair price for this item given the market, the present time of year, demand, and the product's characteristics? Because these variables are always changing, this question is extremely difficult to answer correctly.

**1.1.1**

They can alter in a matter of minutes depending on the goods, which is particularly true in the eCommerce sector. This is why, on a daily basis, companies such as Amazon adjust their product prices millions of times. It is, however, extremely difficult for smaller merchants to compete with Amazon. They have no choice but to compromise; either they don't consider many variables in order to change the prices, or they don't consider many elements in order to change the prices. in a timely manner, or take into account as many elements as possible and hope that the market hasn't changed significantly by the time prices are determined. Pricing methods that have been around for a long time rely entirely on human judgment. The diagram below depicts the basic distinction between human-supervised ML-driven pricing and pricing that is entirely human-powered.

# LITERATURE REVIEW

Sales forecasting is among the fundamental inputs for planning decisions throughout the supply chain. Estimating future demand more accurately is critical for meeting it, while minimising inventory and other related costs. These demand estimates are often modelled based on historical patterns in the data. However, including external information can improve the sales forecast performance (Currie and Rowley, 2010), especially in volatile environments. Earlier work has looked at including additional information from within the supply chain, see for example Aviv (2001), Trapero et al. (2012) and Williams et al. (2014); as well as price and promotional data (Huang et al., 2014, Ma et al., 2015). Bertrand et al. (2015) discusses the importance of weather information in retail sales. The main focus of this stream of research has been improving operational forecasts. In contrast, tactical level dynamics can be different due to the nature of planning, the relevant horizons and the business models. Macroeconomic indicators can contain leading context information, such as changing global economic conditions. Companies review their national markets looking at the evolution and future expectations of economic indicators. These leading indicators are typically published on monthly or lower frequency, making them too slow for forecasting for operational purposes. However, for medium to long-term horizons, these macroeconomic indicators could enrich the forecasts. In several sectors tactical forecasting that supports plans for raw materials, labour, machine resources and financial planning, has a horizon of 3 to 12 months ahead. In this context macroeconomic information is relevant. Often tactical level forecasts rely on univariate methods, which are unable to model changing conditions in a market. That forces organisations to rely on expert adjustments for this purpose, which are characterised by various biases and being unstructured (Fildes et al., 2009). In contrast to a fully statistical approach, this human interaction increases the complexity 2 of the forecasting process and severely limits the extent to which it can be automated

### Literature Review Summary

Table 2.1: Literature review summary

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Year and citation** | **Article Title** | **Purpose of study** | **Tools/ Software used** | **Comparison of technique done** | **Source**  **(Journal/**  **Conference)** | **Findings** | **Data set (if used)** | **Evaluation parameters** |
| 2018 | Machine Learning for Retail Price Recommendation with Python | To build a model that automatically suggests the proper product prices. | * Python 2.7 or higher * Kaggle * Google Collab | exploit purchase data to predict inventory needs in real-time. | https://towardsdatascience.com/machine-learning-for-retail-price-suggestion-with-python-64531e64186d | automatically suggested the product price | train.csv | * Accuracy * Complexity |
| 2021 | How to Build a Price Recommender App with Python | To take a glance at the simplistic price optimization approach and also build a simulator app. | * Python 2.7 or higher * Kaggle * Google Collab | accurately predict how customers will react to certain prices and forecast demand for a given product. | https://www.analyticsvidhya.com/blog/2021/08/build-a-price-recommender-app-with-python/ | pricing optimization  app | Price.csv | * Accuracy * Complexity |

# PROBLEM FORMULATION

Predicting the price of a product is a tough challenge since very similar products having minute differences such as different brand names, additional specifications, quality, demand of the product, etc. can have very different prices. For example, one of these sweaters cost $335 and the other cost $9.99. Can you guess which one’s which?

Price prediction gets even more difficult when there is a huge range of products, which is common with most of online shopping platforms. It’s highly challenging to predict the price of almost anything that is listed on online platforms.

# OBJECTIVES

The proposed work is aimed to carry out work leading to the development of an approach for the Retail Price Prediction Model. The proposed aim will be achieved by dividing the work into the following objectives:

1. Minimizes the risk usually involved in changing prices thanks to its prediction capabilities.

2. Retail teams can essentially use machine learning to test out various promotions or pricing strategies to understand what their impact may be.

3. To make the pricing decisions of pricing managers more profitable.

4. **Predict how customers will react to certain prices and forecast demand for a given product.**

# METHODOLOGY

The following methodology will be followed to achieve the objectives defined for the proposed research work:

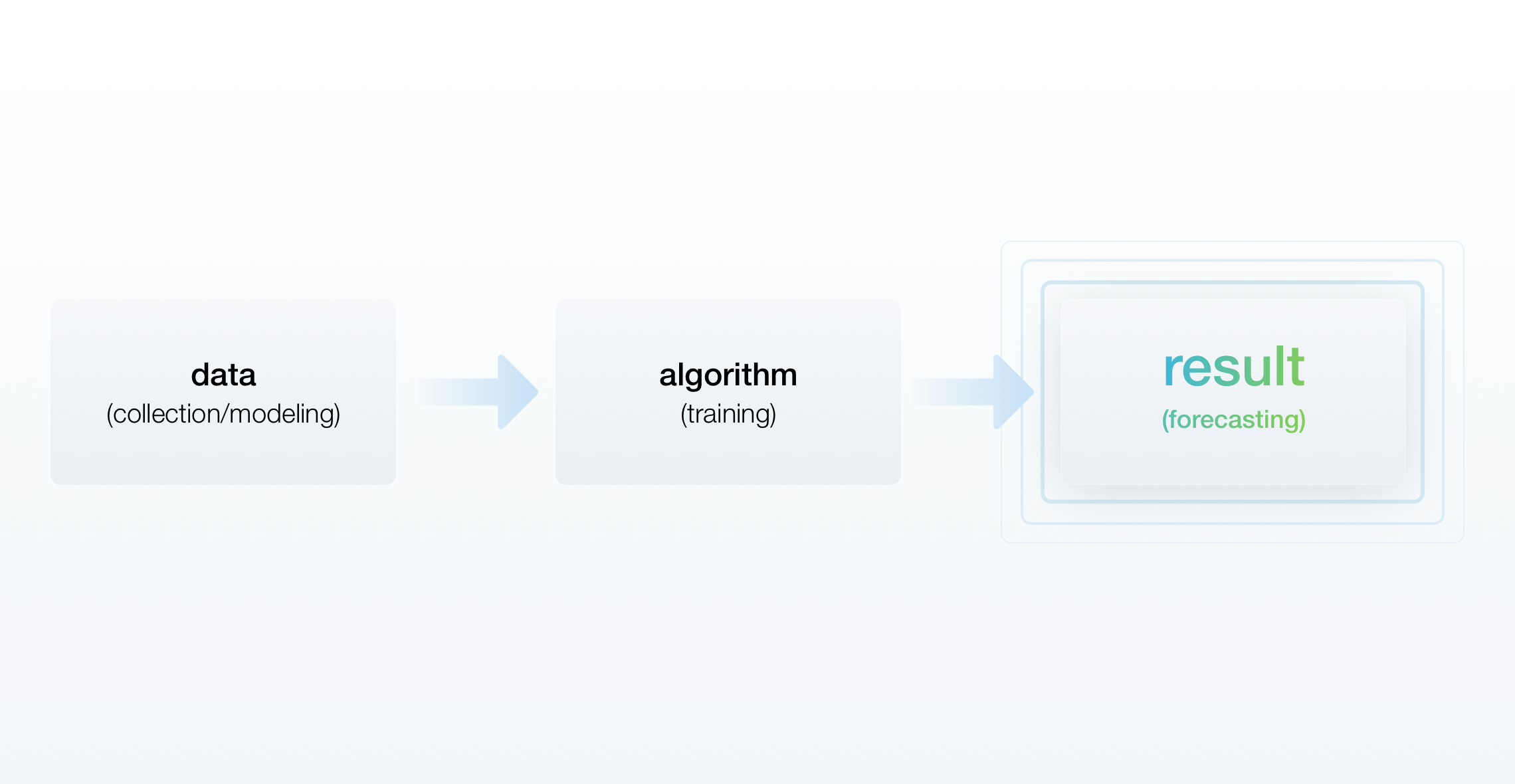
1. Detailed study of **Retail Price Prediction** will be done.
2. Installation and hands-on experience on existing approaches of **Machine learning** will

be done. Relative pros and cons will be identified.

1. Various parameters will be identified to evaluate the proposed system.
2. Comparison of newly implemented approaches with exiting approaches will be done.

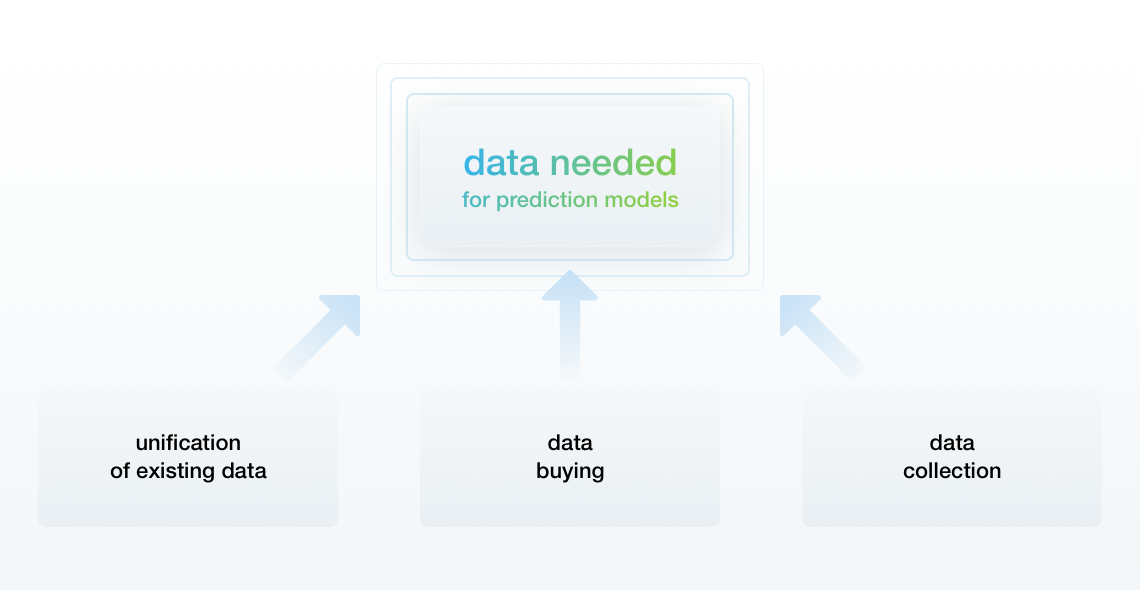
Algorithms must learn from historical and competitive data before being deployed. During the learning step, the model (or algorithm) examines every single variable that affects sales, such as pricing and traffic. The model is ready for a pilot and if the merchant is satisfied with the results, for continued use once the training is completed and the algorithm provides correct predictions that are later confirmed by real results.

Retailers frequently have incomplete, difficult-to-extract data or data that is ill-structured and in an improper format. We'll go through how machine learning deals with insufficient data in retail in the sections below.

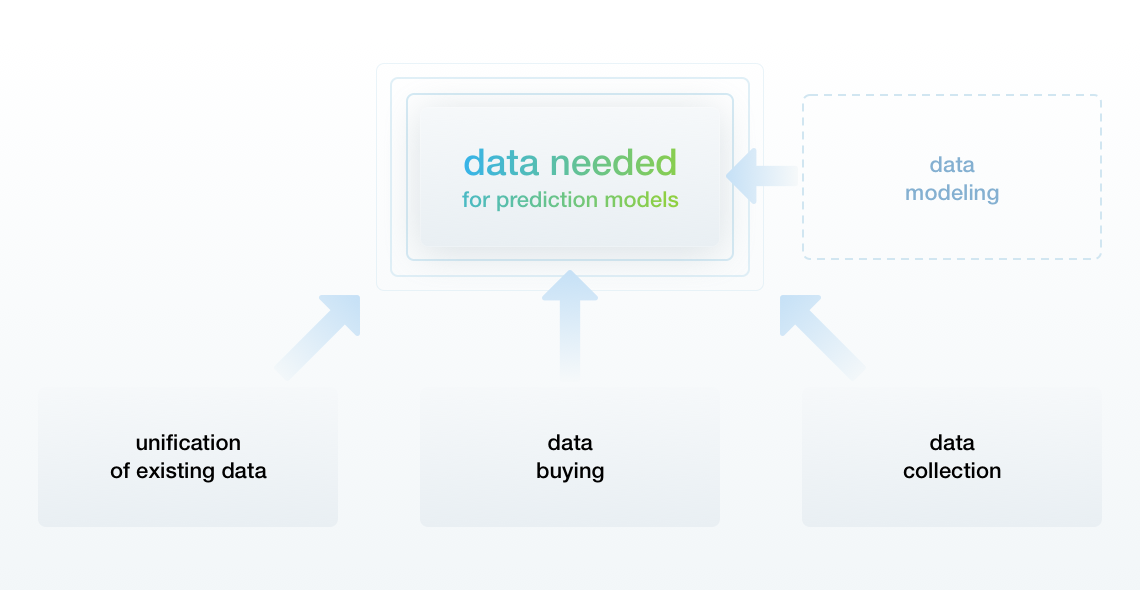


Retailers must combine all data into a single format. In addition, if a shop has previously acquired some data and then adds new data based on other criteria, such as competitive prices, the company must wait nearly a year to begin collecting new data.

Another option is to buy the missing information.



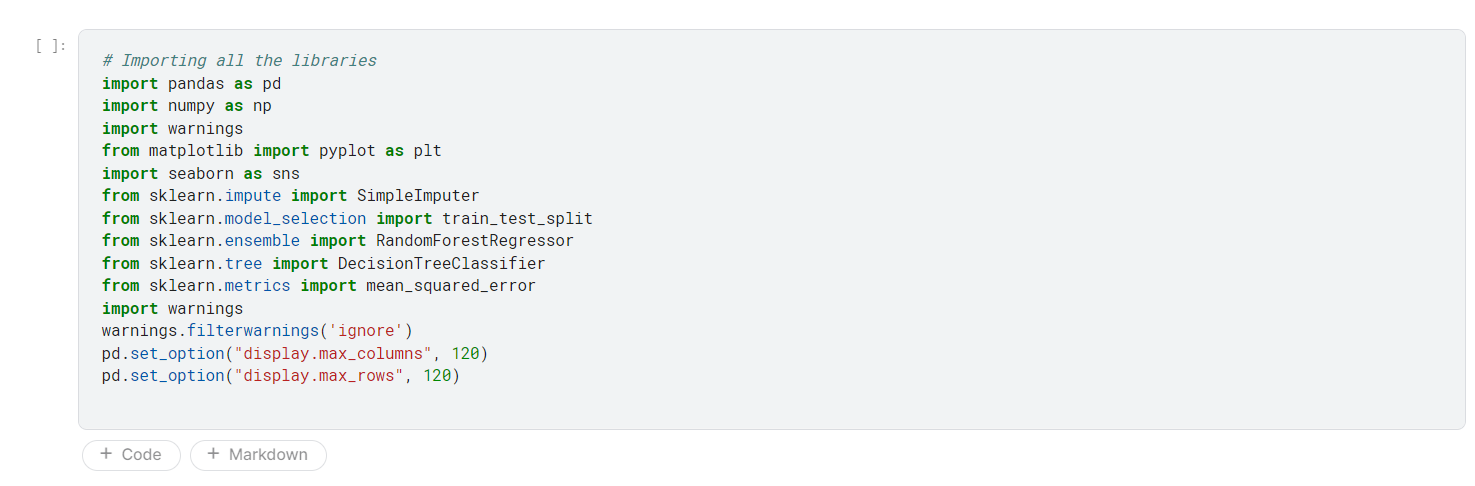
If there is no way to obtain the necessary data, the algorithms can use data modeling methods to simulate it

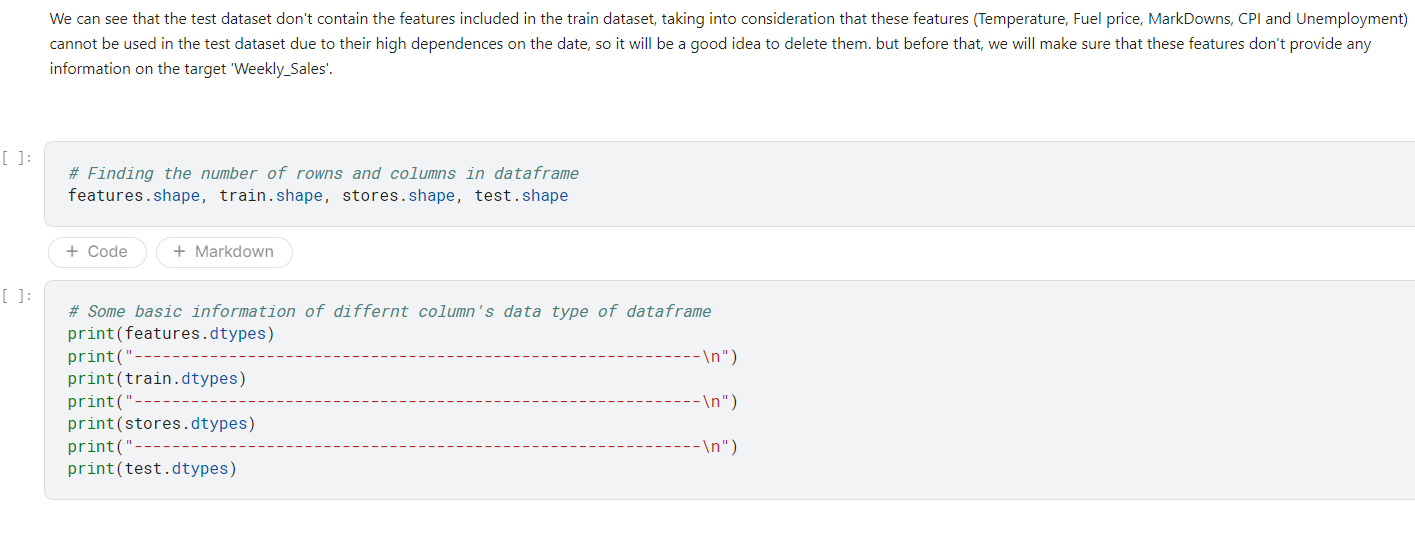


**6 PROGRESS**

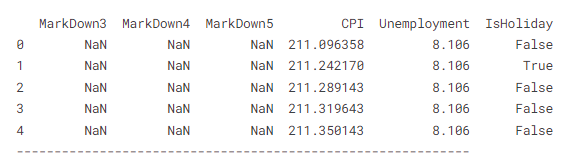
1. Exploratory Data Analysis to describe and clean the data, and to understand attributes
2. Feature selection to keep only important attributes
3. Developing a framework to evaluate and spot-check algorithms

In addition, Retail runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labor Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks. the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data.

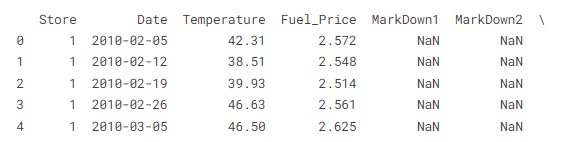




Further the code for Is Holiday has been prepared we can enter the holidays of a month in the web UI in later development of the software.

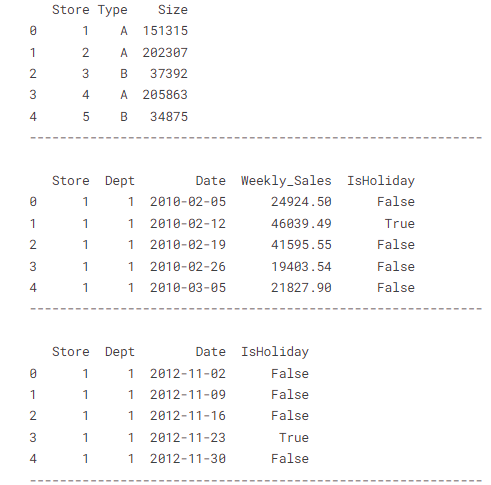


The function for predicting the outcome using the variables like temperature and fuel price has been added to enhance the functionality of the software



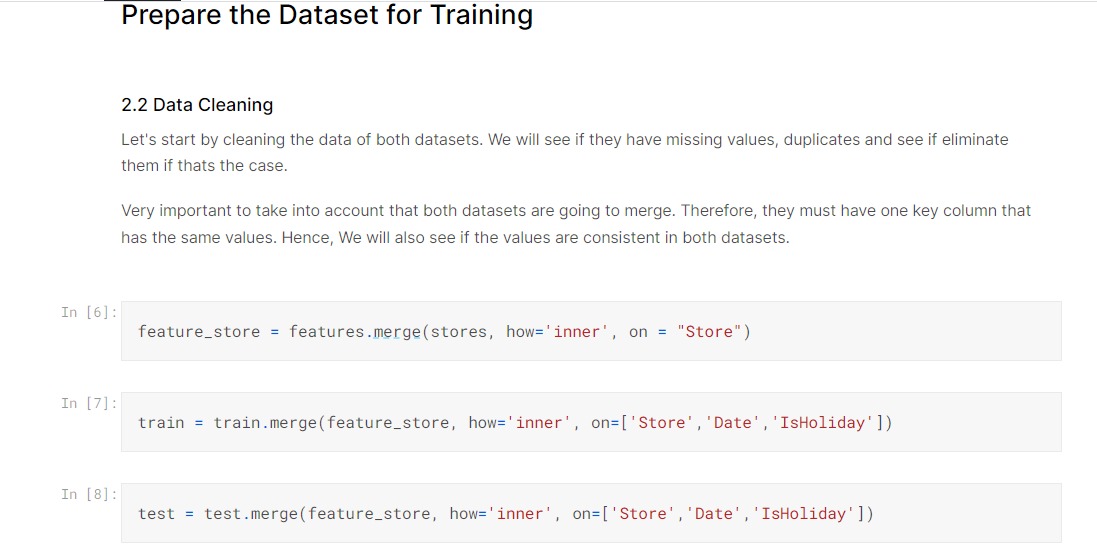
The Ongoing development in that software is to add some of the outlet sizes and specifying the the type of store.

And then we can classify the stores on the basis of the size of the store, this will greatly impact the prediction as the storage for the goods is and essential variable that can be very crucial in making the sales high.

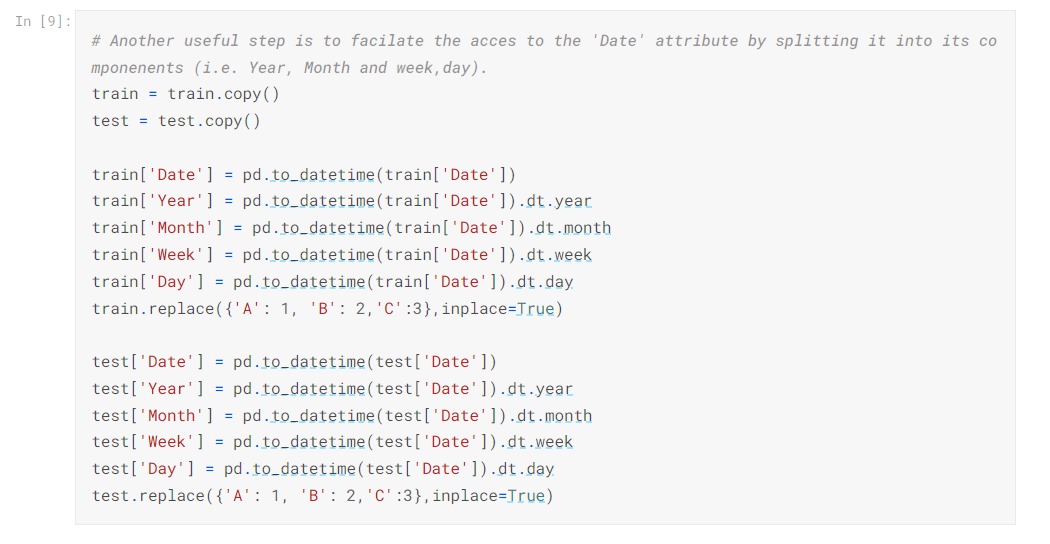


The following code has been done to verify the dataset provided and making the data more precise by cleaning the data.

The data has been prepared by following code:-

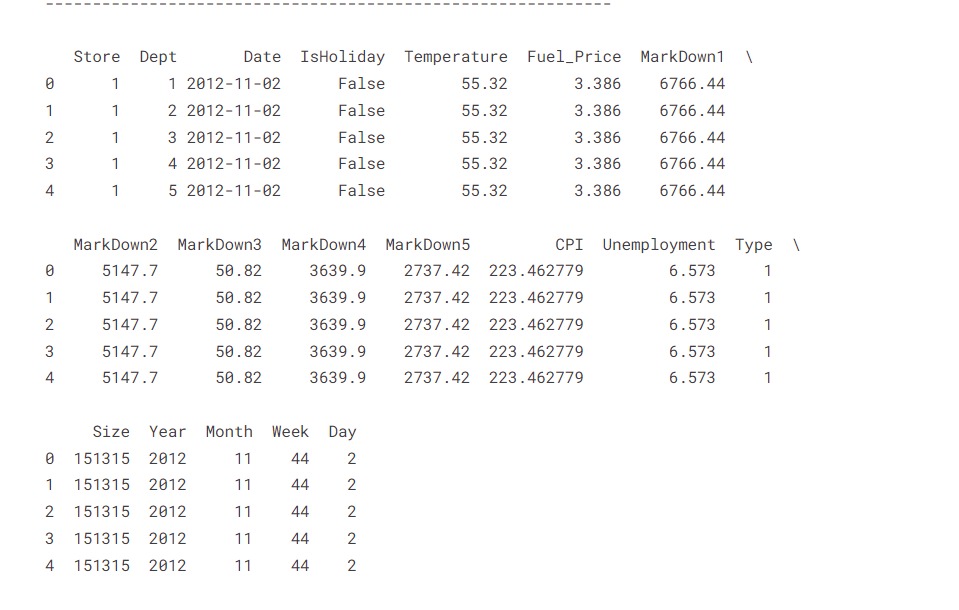


Another useful step is to facilitates the access to the Date attribute by splitting it into components I.e. Year, Month, and week, days



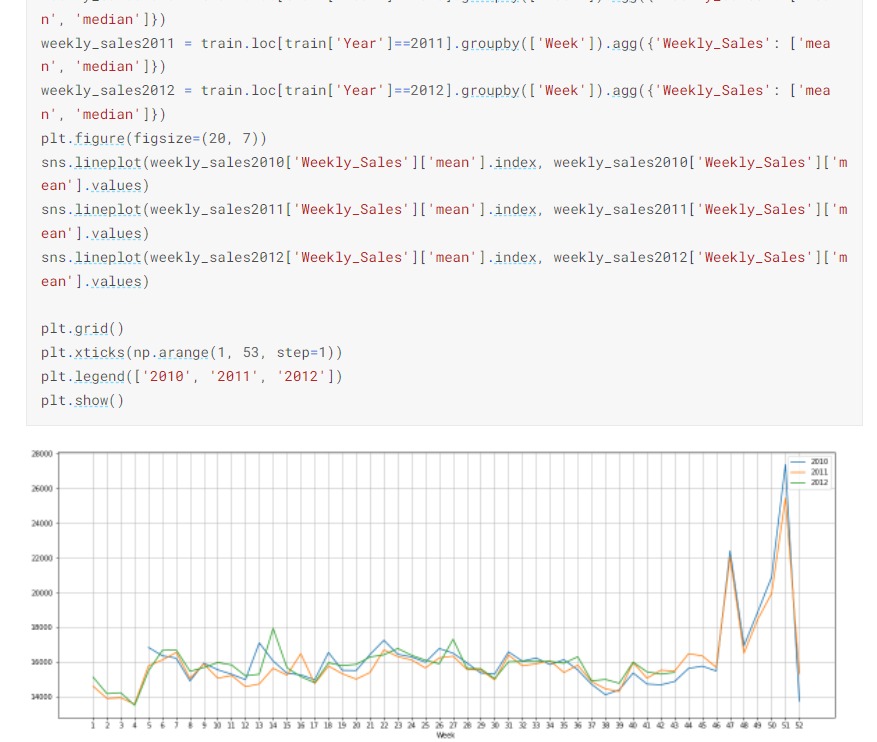
The following output has been received on running the above code part The attributes like is holiday, temperature, Fuel Price and Markdown price has been considered.

After that the size of the stores and the attributes like Year, Month, Week and days were considered

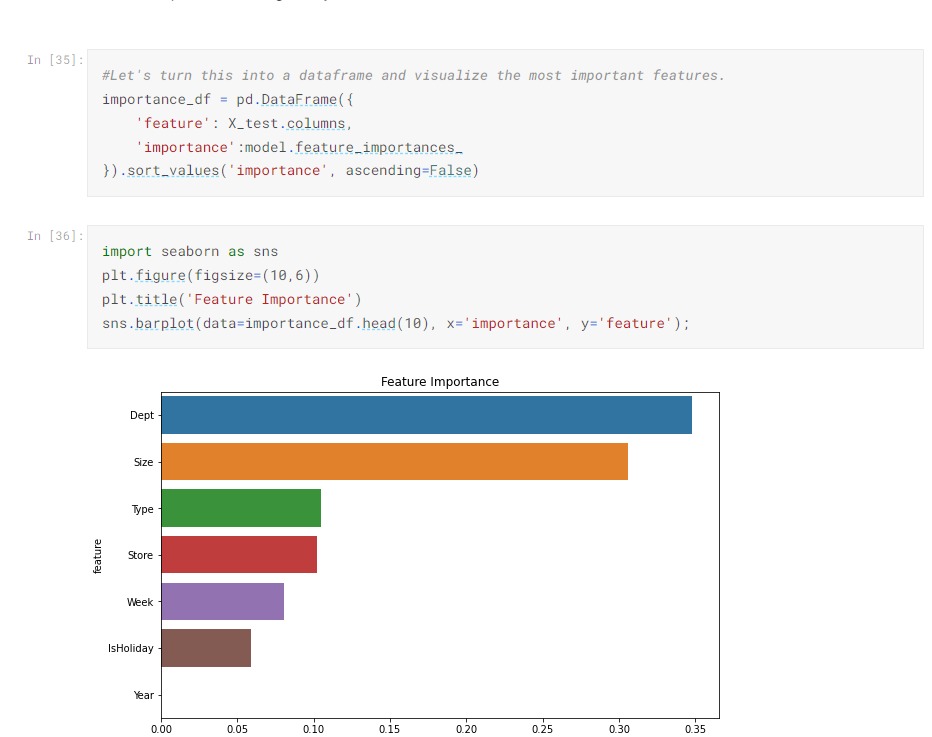


Then the graph and plot for the same has been ploted keeping in mind that we have to maximize the profits based on the various attributes and where ever it will show peaks in the graph we will make sales of the retails on that day.

As that will be the most benificial day for the retail sales and the sales and profit will be maximized



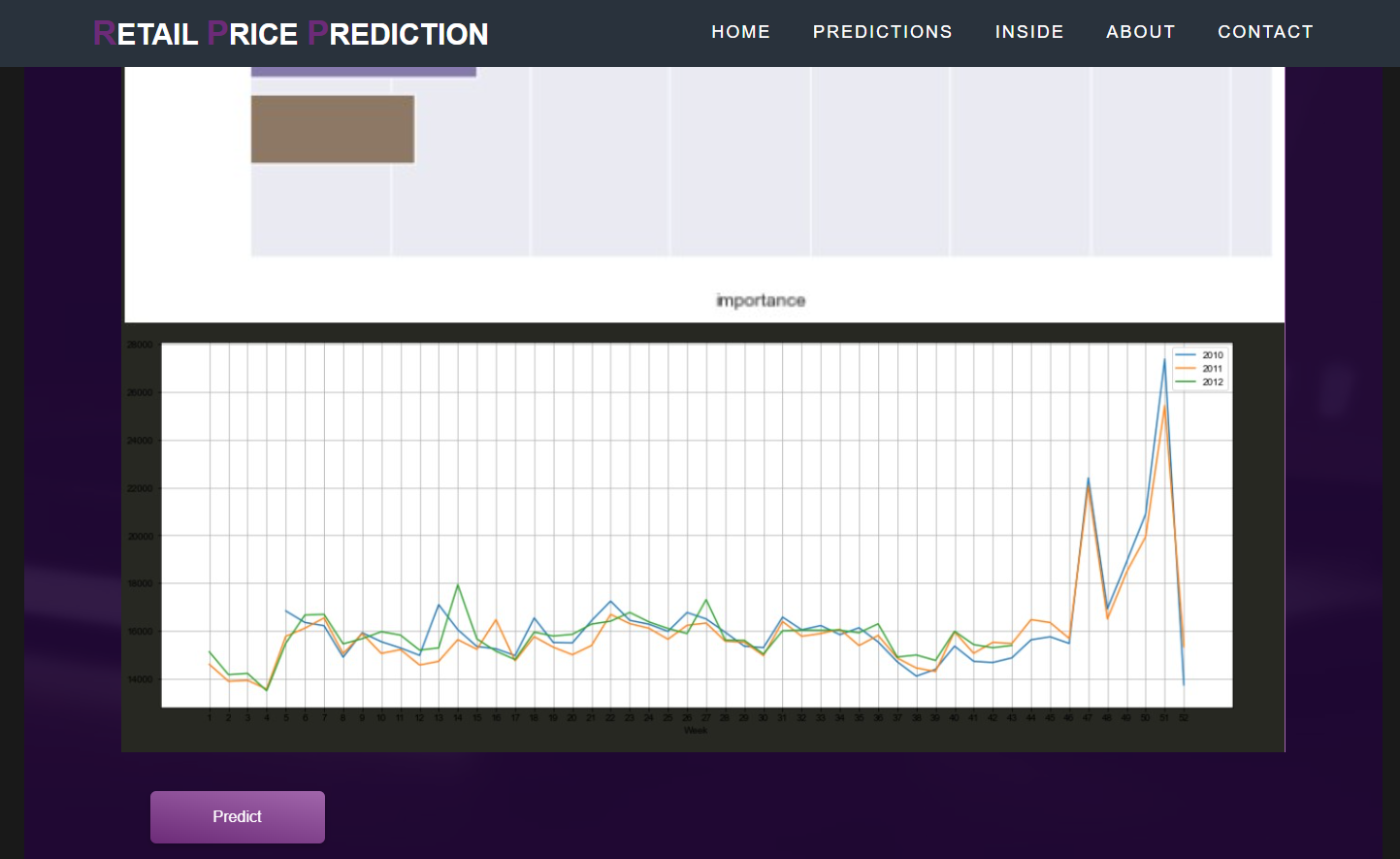
We will later analyze the features that are mostly important in predicting the optimal price for the goods to be sold in a whole month

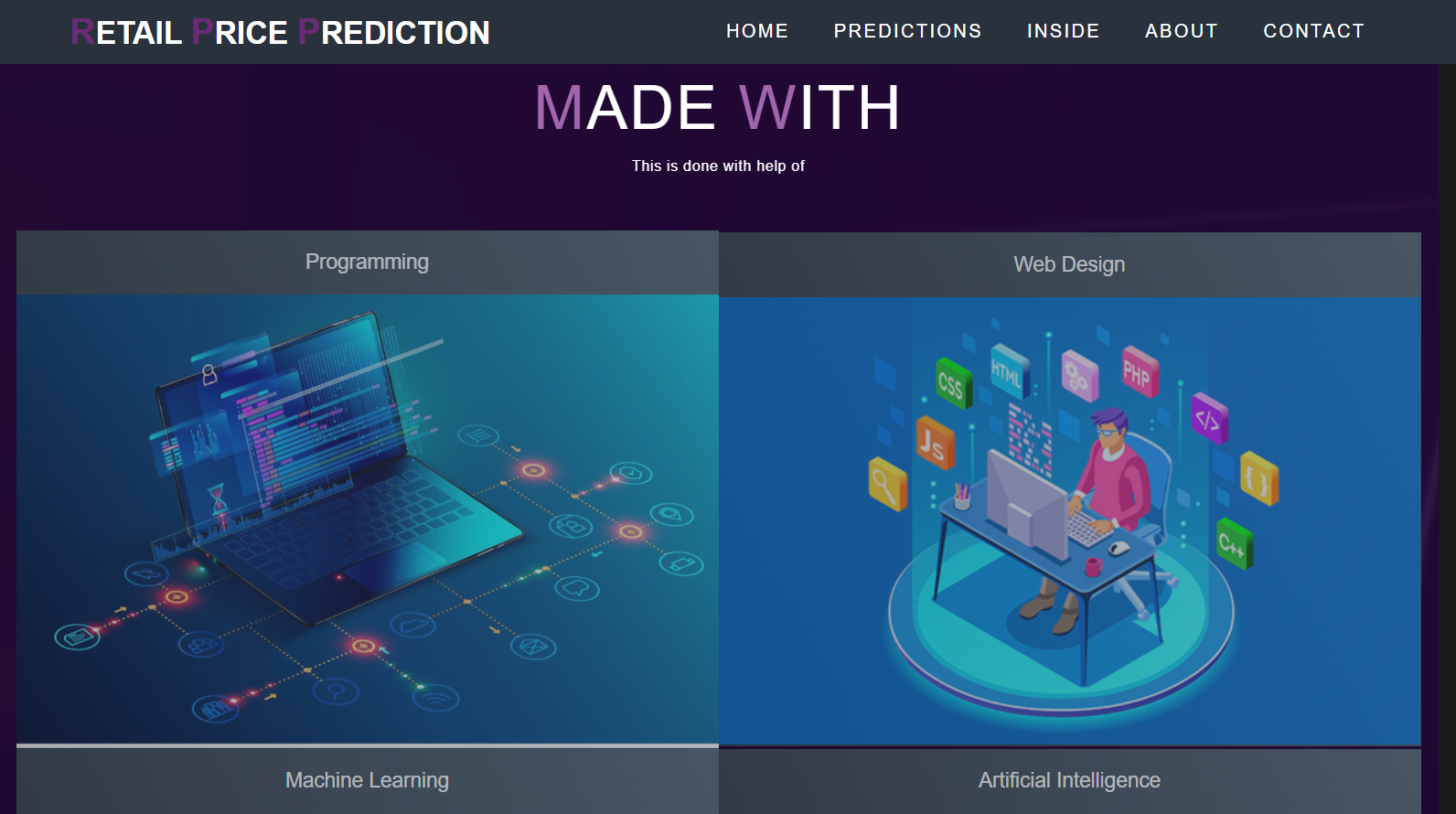


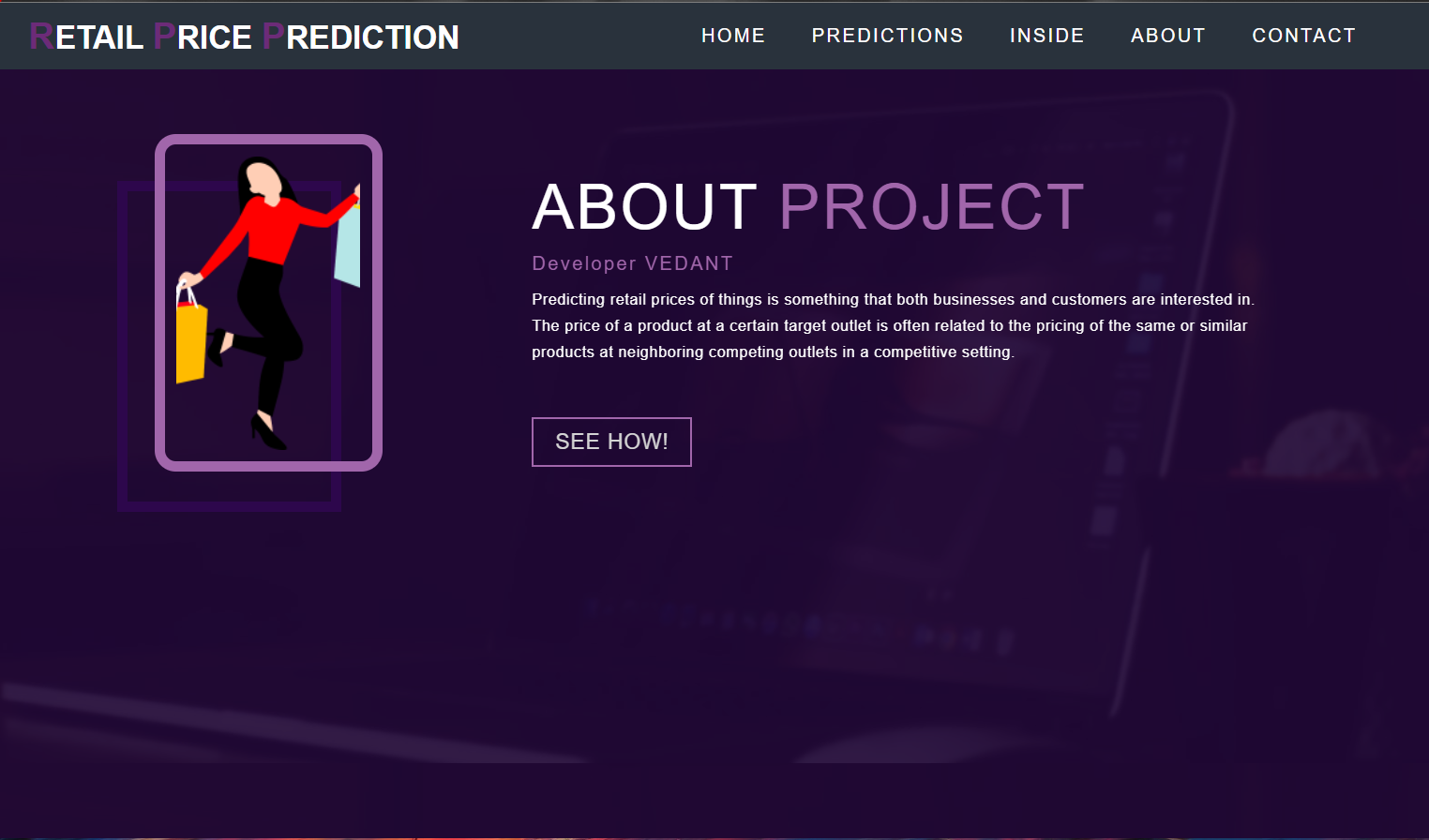
**7 WEB UI**

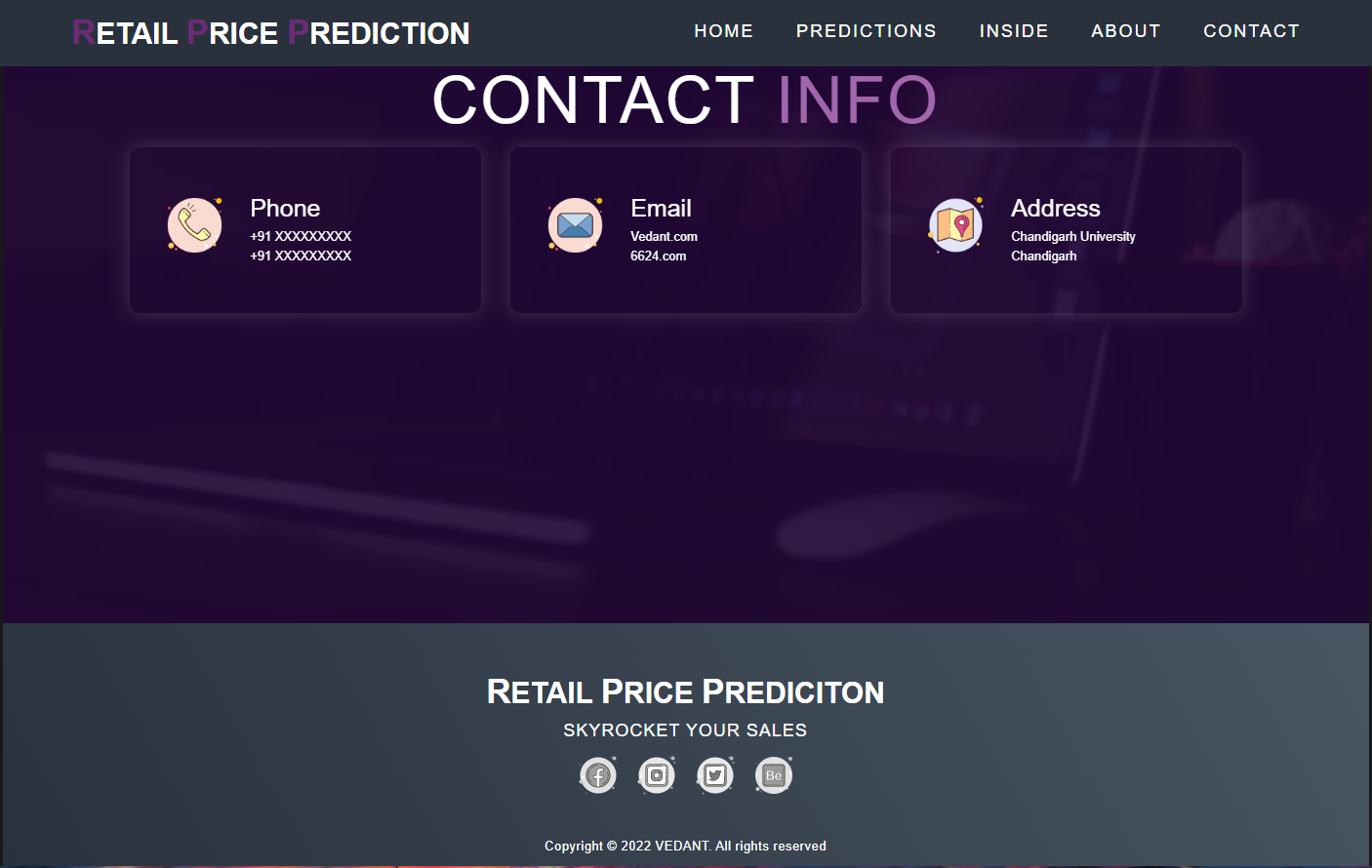












**8 TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK**

**CHAPTER 1: INTRODUCTION**

This chapter will cover the overview of the Retail Price Prediction Model.

**CHAPTER 2: LITERATURE REVIEW**

This chapter includes the literature available for the Retail Price Prediction Model. The findings of the

researchers will be highlighted which will become the basis of the current implementation.

**CHAPTER 2: BACKGROUND OF PROPOSED METHOD**

This chapter will provide an introduction to the concepts which are necessary to understand the proposed system.

**CHAPTER 4: METHODOLOGY**

This chapter will cover the technical details of the proposed approach.

**PUBLICATIONS (Optional)**

**REFERENCES**

# REFERENCES

.

1. J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio. Theano: [A cpu and gpu math compiler in python](https://projects.scipy.org/scipy2010/slides/james_bergstra_theano.pdf). In Proc. 9th Python in Science Conf, pages 1–7, 2010
2. M. Y. Park and T. Hastie. [L1-regularization path algorithm for generalized linear models](http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2007.00607.x/abstract;jsessionid=315EA2A9E59E2A7E79041A04F4FE065D.f02t01?userIsAuthenticated=false&deniedAccessCustomisedMessage=). Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(4):659–677,