Fundamentals of Data Mining - IT3051

Group Project – G05

Statement Of Work Document

IT Number	Student Name
IT20165352	Peiris M.I.M
IT20074340	Nirmal M.D.S
IT20101510	Alwis W.C.H
IT20619558	Samarakkody G.K.G.D.A
IT20201364	Thathsarani R.P.H.S.R

Date of Submit

15/10/2022

Table of Contents

Background

For the following assignment , the group will be focusing on a problem which is faced by many people who are engaging in their day-to-day business ,academic , and professional work with the use of Technology Infrastructure.

With the rising component costs, increased supply chain shortages, logistic costs resulting from the pandemic and the energy crisis all over the world. The unusual hike of the prices in the prevailing technology market has proven to be a difficult time for the information technology workforce and the students in the information technology industry in choosing the laptop which matches their requirement as well as fitting their budget.

Various models of computers are sold in the market at different prices. The price of the computer will be mainly based on some key features like RAM, CPU, GPU ,memory ,screen size and resolution, type name, operating system, and brand(company) of the laptop.

As a solution for this business problem, a group of Data Science students has decided to build a web application that integrated by a machine learning model which will be able to make a prediction of laptops' prices based on their key features. Web application will facilitate people in choosing the components they need , and they could obtain an idea about what will be the expected price of the laptop they want. And see whether that laptop fits their budget. And if not, they are able to change or adjust the requirements freely and find the laptop which matches exactly to their budget and the requirements they wish to have.

Not only predicting the price of laptop but also the proposed solution will allow users to choose their computer from anywhere, and in few minutes effortlessly.

To conclude , as data science team our endeavor is to obtaining high accuracy for "Laptop Price Prediction Model" , thereby contribute to information technology sector to offer better solution for above mentioned term problem.

Scope of work

The main objective of this project is to provide a data mining solution for the above-mentioned real-world problem. The problem is related to predicting laptop prices according to the requirements of the user who needs to buy laptops. Prior analysis of the data set revealed the variables and how they impact the prices of laptops. Therefore, regression is the reliable method which use as the data mining function to build models after applying specific algorithms. The optimal model owes the higher accuracy which is going to deploy as a web application.

As the final output, the optimal model will be developed and deployed as a single convenient web application using web development tools. The user possesses the ability to provide the required features and based on the user's requirements the model will predict the relevant prices. According to the timeline, the project will process and the time duration for every task of the project plan is mentioned below. After the deployment of the web application final report will be submitted at the end.

Activities

Activity No	Activity
01	Select a domain, define the problem statement, and identify dataset relevant to the problem.
02	Inspect the selected dataset and select appropriate data mining techniques as per its nature.
03	Segregate the tasks according to the selected technique and allocate them for the group members.
04	Select suitable tools and technologies to perform the required tasks.
05	**Regression Technique**
a.	Study and understand the dataset.
b.	Data preprocessing and preparation.
c.	Build the models based on algorithms selected for the respective technique.
d.	Evaluate the algorithm performance based on its accuracy and other metrics.
06	Design user interfaces for the web application and develop final application along with the data mining models.
07	Deploy the application and perform testing based on test cases defined.
08	Create a report consisting of all the key information on the project, its process and final output.
09	Create a video to demonstrate the work to users and interested stakeholders.

Approach

Purpose	Tools/Methodology
Development Methodology	Agile Methodology
Dataset	From Kaggle Link:- https://www.kaggle.com/datasets/aggle6666/laptop-price-prediction-dataset/code
Data Mining Technique & Algorithms	Regression Liner Regression, Random Forest, Decision tree regressor
Programming Language	Python
Frontend	HTML/CSS(Bootstrap)
Backend	Flask
Server/Hosting	Heroku

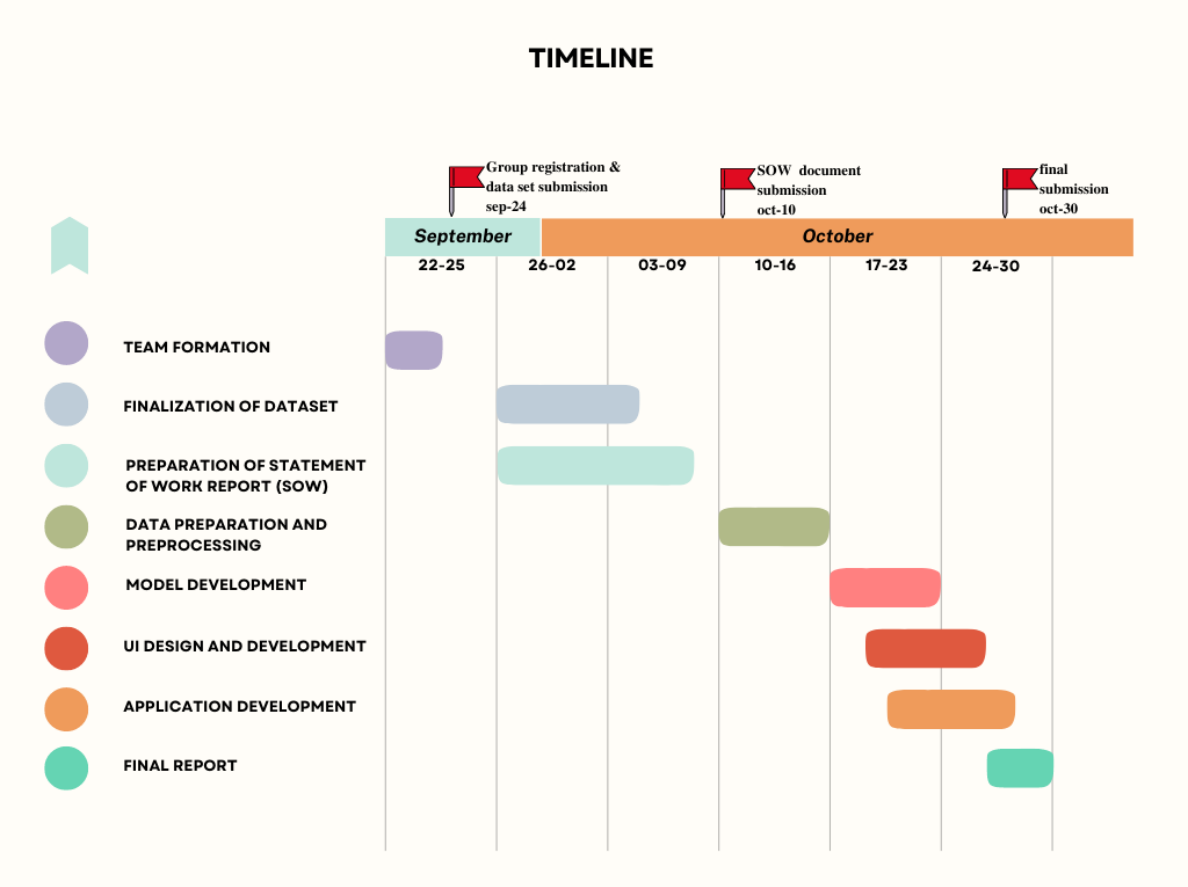
Note : Technologies and algorithms are subject to change.

Deliverables

Delivery	Description
Statement of Work (SOW) Report	The SOW document includes key information about problem background, scope of work, activities, approach, deliverables, project plan and timeline, assumptions, individual roles and responsibilities.
Data Mining Models	According to problem domain, final implementation would be a predictive output (solution) for the problem. Supervised machine learning technique is used for building the model.
Web Application	Web application integrated with machine learning model to make user operations.
Video Presentation	Demonstration of the features of the application
Final Report	The report consists of information on project domain, solution developed and test cases.

Project Plan & Timeline

Type	Title	Start date	End date	Duration	% Complete
Task	Team Formation	09/22/2022	09/25/2022	3	100
Task	Finalization of Scope & Datasets	09/26/2022	10/06/2022	7	100
Milestone	Group Registration & Dataset Submission	09/26/2022	09/26/2022	0	100
Task	Preparation of Statement of Work (SOW) Report	09/26/2022	10/08/2022	7	100
Milestone	Approval on Statement of Work	10/09/2022	10/09/2022	-	100
Milestone	Submit Statement of Work	10/10/2022	10/15/2022	5	-
Task	Data Preparation & Preprocessing	10/11/2022	10/17/2022	-	0
Milestone	Finalization of Tools & Models	10/15/2022	10/16/2022	-	-
Task	Model Development	10/17/2022	10/23/2022	-	0
Task	User Interface Design & Development	10/20/2022	10/25/2022	-	0
Task	Finalization & Application Deployment	10/26/2022	10/27/2022	-	0
Task	Video Presentation	10/28/2022	10/29/2022		
Task	Final Report	10/27/2022	10/30/2022	-	0
Milestone	Submission of Report, Application	10/30/2022	10/30/2022	-	-



Assumptions

- The data is based on an old dataset, but we assume that the prices are from recent years.
- The price of the laptops is in Euros. We can convert it into any currency as we want.
- We assume the users of this system include students and professionals who has basic understanding on how to choose components for their laptop according to their requirement.

Project team, roles, and responsibilities

Member	Role	Responsibilities
Peiris M.I.M	Team Lead	• Selection of problem domain. • Mine and analyze the data. • Data visualization.
Nirmal M.D.S	Integrator	• Apply data into models. • Build the models. • Integrate the data mining model to the developed application.
Alwis W.C.H	Technical Expert	• Selection of technology and tools. • Deploying the web service. • Design the website. • Build the models.
Samarakkody G.K.G.D.A	Domain Expert	• Selection of ideal data mining models for the project. • Enhancing accuracy of the models. • Data preprocessing and preparation.
Thathsarani R.P.H.S.R	Integrator	• Data preprocessing and preparation. • Integrate the data mining model to the developed application. • Testing the data mining model.

References

- https://analyticsindiamag.com/top-6-regression-algorithms-used-data-mining-applications-industry/
- https://www.javatpoint.com/regression-analysis-in-machine-learning
- https://www.jigsawacademy.com/popular-regression-algorithms-ml/
- https://towardsdatascience.com/organizing-machine-learning-projects-e4f86f9fdd9c
- Understanding Machine Learning: From Theory to Algorithms (Shai Shalev-Shwartz and Shai Ben-David)
- Han, J., Kamber, M., and Pei, J., Data Mining: Concepts and Techniques, 3rd Ed., Elsevier Inc., 2012.