Helwan University

Project Python

Predict Price phone

Project Team:

- Khaled Ayman Farouk
- Emad Rabie Hasb elnaby
- Omar Hussein Abd EL Rady
- Ahmed Shaban Ragab
- Mohamed Ahmed Mohamed

Introduction:

Our project, a Phone Price Prediction application, designed to assist individuals in determining the value of their mobile phones when buying or selling. With the ever-evolving smartphone market and the potential for deceptive sellers, our application aims to provide users with accurate price estimates based on the phone's specifications. By inputting key features such as RAM, battery, storage, front camera, screen size, and operating system, users can obtain a range of prices, thus empowering them to make informed decisions and avoid potential scams.

User Interface:

The application leverages the intuitive Streamlit framework to create a user-friendly interface. Users are presented with input fields to enter the phone's specifications, allowing for a seamless experience.

Feature Input:

Users can input various phone capabilities, including:

- RAM: The amount of memory in gigabytes (GB)
- Battery: The battery capacity in milliampere-hours (mAh)
- Storage: The storage capacity in gigabytes (GB)
- Front Camera: The front camera resolution in megapixels (MP)
- Screen Size: The screen size in inches
- Operating System: The phone's operating system, such as Android, iOS, or other

Process on the data

The topic of predicting phone prices using web scraping, data cleaning, machine learning, data visualization, and Streamlit.
Highlight the importance of accurately predicting phone prices for various stakeholders (consumers, retailers, manufacturers).

Web Scraping:

The Magic of Web Scraping:

Web scraping has revolutionized the way we gather data from the internet. It provides us with a powerful mechanism to extract valuable information from websites, including product details, customer reviews, and, in our case, phone prices. By automating the data collection process, we can access vast amounts of data quickly and efficiently.

Web Scraping from Amazon:

Amazon is renowned for its extensive product offerings across various categories, including mobile phones. We recognized Amazon's credibility and comprehensive range of phone models and associated pricing information, making it an ideal choice for our web scraping process. By scraping data from Amazon, we were able to collect up-to-date and real price data, ensuring the accuracy of our price prediction system.

Extracting Phone Specifications:

During the web scraping process, we extracted essential phone specifications such as RAM, battery, storage, front camera, screen size, and operating system. By harnessing the power of Amazon's product pages, we gathered a diverse range of phone models, covering different brands, categories, and price ranges. This comprehensive dataset allowed us to create a robust and accurate price prediction model.

Data cleaning:

Importance of data cleaning:

Data cleaning is a crucial step in preparing collected data for analysis and modeling. It involves removing raw and null values, as well as eliminating irrelevant or inconsistent data. Addressing data inconsistencies, such as handling missing values, standardizing units of measurement, and resolving formatting inconsistencies, ensures data quality and reliability.

Handling Missing Values:

Delete rows or columns with prevalent missing values or impute values based on available data.

Removing Raw and Null Values:

Eliminate data that is not in a usable format or contains null values.

Eliminating Irrelevant or Inconsistent Data:

Remove duplicate records, erroneous entries, outliers, or data that does not align with research objectives.

Standardizing Units of Measurement:

Ensure consistency and comparability by converting units, scaling values, or establishing a common reference point.

Resolving Formatting Inconsistencies:

Address variations in date formats, numerical representations, or text case.
Data cleaning enhances data quality, minimizes biased or erroneous insights, improves model accuracy, and ensures reliable analysis.

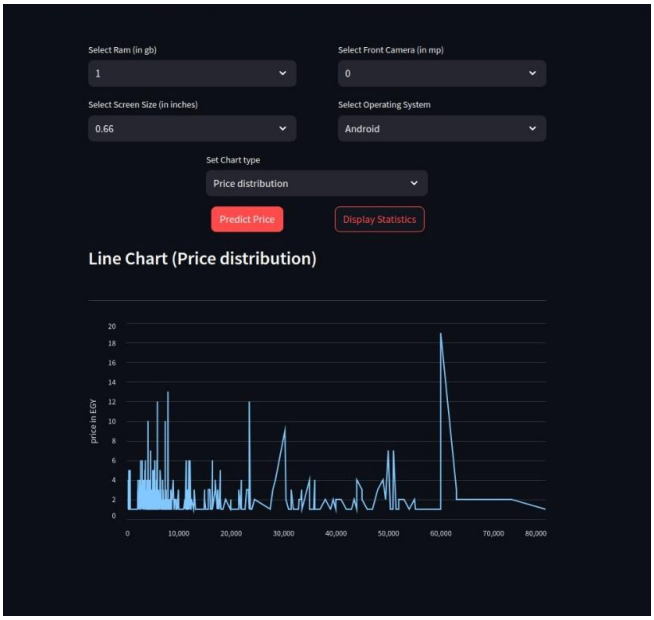
Data Visualization:

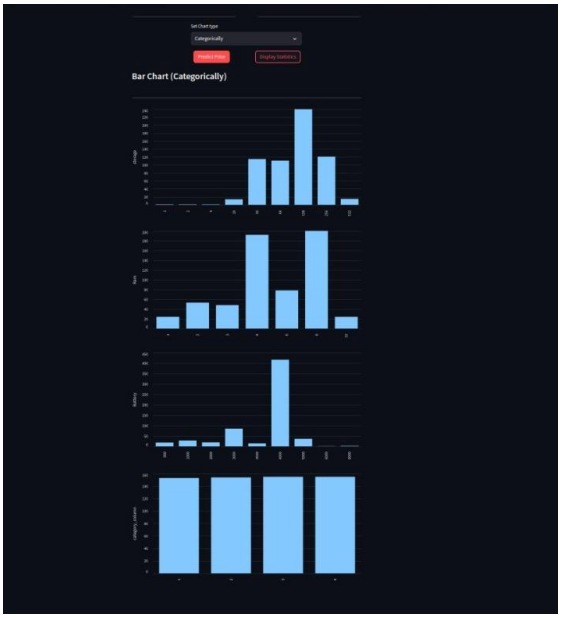
Data Visualization:

Visualization plays a crucial role in understanding the dataset and identifying relationships between variables. We employed various data visualization techniques using libraries such as Matplotlib and Seaborn. Our visualizations highlight the correlation between phone features and prices, allowing us to identify which features have the most significant impact on pricing.

Categorizing Price Ranges:

To enhance the accuracy of our predictions, we categorized the price data into four distinct ranges. This categorization enables us to train our machine learning model more effectively and provides users with a clearer understanding of where a specific phone falls within the price spectrum.





Machine learning:

During the machine learning phase, we utilized the train_test_split function from the sklearn.model_selection module to split our dataset into training and testing sets. This division allowed us to evaluate the performance of our model on unseen data.

Dataset Split:

By randomly splitting the dataset, we ensured that the model learns patterns from the training set and generalizes well to new, unseen data. The training set was used to train the model, while the testing set served as an independent evaluation set to assess its performance.

Algorithm Selection:

For our classification task of predicting phone price categories, we employed the Logistic Regression algorithm (LogisticRegression). Logistic Regression is a well-established algorithm that is effective in handling binary classification problems like ours. It learns a decision boundary that separates the data into different classes based on input and output price categories features .

Accuracy:

The accuracy after train and test the model is 75%.

Saving the Model:

To ensure seamless integration and reusability, we utilized the pickle library to save the trained model, enabling us to load and use it in other applications, such as our Streamlit application.

Web Application:

Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. It is a Python-based library specifically designed for machine learning engineers,In order to provide a user-friendly experience, we developed a Streamlit application. Users can input phone specifications, including RAM, battery, storage,inches, front camera, screen size, and operating system (Android, iOS, or other). The application then displays the predicted price range and recommends phones within the specified range, based on the input features.