**Laptop Price Prediction using Machine Learning: A Comparative Study of Regression Models**

* SAGNIK SAMANTA

**Abstract** In today's rapidly evolving technological environment laptops are essential tools for individuals and businesses alike. Accurate prediction of laptop prices helps retailers develop effective pricing strategies and enables consumers to plan their budgets and select the most suitable laptops. This study explores multiple machine learning techniques to predict laptop prices based on key specifications. We employed Multiple Linear Regression, Random Forest Regressor, and XGBoost models to evaluate their performance using Root Mean Squared Error (RMSE) and Adjusted R-Squared (R²). The results indicate that the XGBoost model offers the best prediction accuracy, outperforming other models in minimizing error and maximizing predictive capability.

**1. Introduction** Laptops have become an essential tool for individuals and businesses, making price prediction an important task in market analysis. This research aims to develop a machine learning model to predict laptop prices based on features such as brand, specifications, and operating system. The study is structured following the standard format of research articles published in Wiley journals, adhering to their guidelines for empirical studies.

**2. Materials and Methods**

**2.1 Data Collection and Description** The dataset consists of 1275 observations with 23 features, including:

* **Numerical Features**: Screen size, RAM, weight, CPU frequency, storage capacity.
* **Categorical Features**: Brand, product type, OS, touchscreen, GPU/CPU company.
* **Target Variable**: Laptop price (in Euros).

**2.2 Data Preprocessing and Feature Engineering** To ensure data consistency, preprocessing steps included:

* Handling missing values.
* Encoding categorical variables using Label Encoding.
* Normalizing numerical features with StandardScaler.
* Splitting data into training (80%) and testing (20%) sets.

**2.3 Machine Learning Models** Three machine learning models were employed:

* **Multiple Linear Regression**: A statistical approach to model relationships between independent variables and price.
* **Random Forest Regressor**: An ensemble learning technique using decision trees for better generalization.
* **XGBoost Regressor**: A gradient boosting technique optimizing predictive performance.

**3. Results and Discussion**

**3.1 Model Evaluation Metrics** Models were evaluated using RMSE and R²:

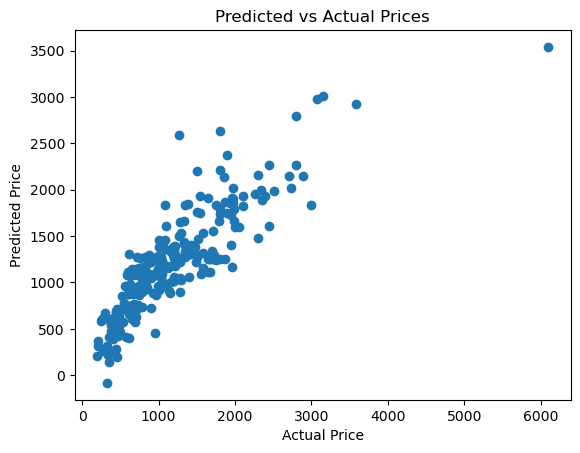
* **Linear Regression**: RMSE = 358.02, R² = 0.74
* **Random Forest Regressor**: RMSE = 251.75, R² = 0.87
* **XGBoost Regressor**: RMSE = 207.79, R² = 0.91

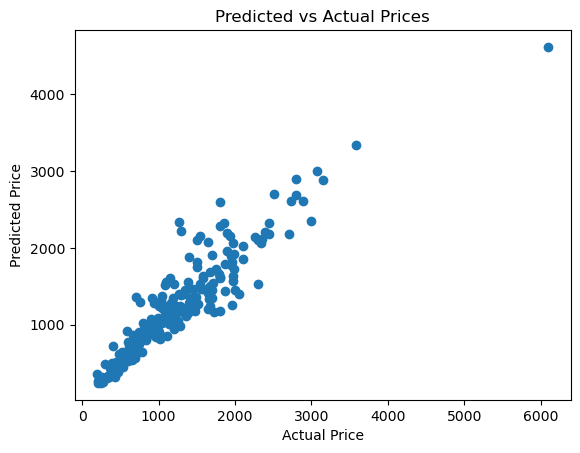
**3.2 Feature Importance Analysis** The XGBoost model demonstrated the best performance, with the lowest RMSE and highest R², indicating a stronger predictive capability. Feature importance analysis revealed that RAM, CPU frequency, and storage type significantly impact laptop prices.

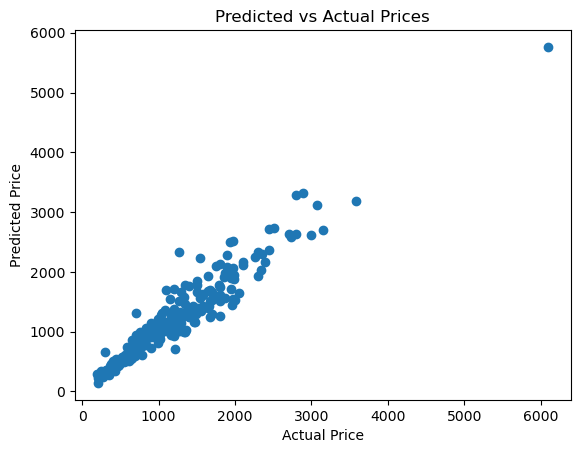
**3.3 Discussion of Results**

* The **Linear Regression model**, while providing a baseline for price prediction, exhibited the highest RMSE, indicating that it struggles to capture nonlinear relationships among variables.
* The **Random Forest Regressor** significantly improved the accuracy, leveraging multiple decision trees to capture complex dependencies between features.
* The **XGBoost model** outperformed both by reducing RMSE and increasing R², proving its efficiency in handling large datasets and complex feature interactions.
* The results align with prior research, emphasizing that ensemble learning techniques yield higher accuracy than traditional regression models.

**3.4 Comparison with Previous Studies** This research aligns with previous studies emphasizing the importance of feature selection and model choice in price prediction accuracy. The findings highlight that ensemble methods like XGBoost provide superior performance over traditional regression models.







**4. Conclusion** This study highlights that advanced machine learning models like XGBoost outperform traditional regression techniques in predicting laptop prices. Future research can incorporate additional factors such as market trends and customer reviews to enhance model performance.

**5. Future Scope** Future research can extend this study in several ways:

* **Integration of Real-time Data**: Incorporating dynamic pricing data and market trends for improved predictions.
* **Sentiment Analysis**: Analysing customer reviews and feedback to understand price perception.
* **Deep Learning Models**: Exploring neural networks to capture complex feature interactions.
* **Feature Expansion**: Including additional specifications such as battery life and build quality.
* **Cross-Market Comparison**: Applying the model to different geographical markets to assess variability in pricing trends.

**6. Acknowledgment** I would like to express my special thanks of gratitude to Mr. Aman Gupta, Data Science Mentor at Unified Mentor for providing guidance throughout this study and helped me to implement the project within the due date. We also acknowledge the contributions of the GeeksforGeeks community for their valuable resources on machine learning. Special thanks to the authors of machine learning books referenced in this study for their insights and methodologies, which greatly influenced our research approach.

**7. References**

1. Kaggle. "Laptop Price Dataset." Available at: [[https://www.kaggle.com](https://www.kaggle.com/)]
2. GeeksforGeeks. "Machine Learning Regression Models." Available at: [[https://www.geeksforgeeks.org](https://www.geeksforgeeks.org/)]
3. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
5. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media.

**Keywords**: Laptop Price Prediction, Machine Learning, XGBoost, Regression Models, Data Analysis.