IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

PREDICTING SMARTPHONE AND LAPTOP MARKET SUCCESS WITHMACHINE LEARNING

[1] P.Gayatri , [2] M.Nandini

[1] Assistant Professor in CST , [1,2] MTech Year

Sanketika Vidya Parishad Engineering College, Visakhapatnam, A.P, India.

Abstract:

This thesis presents a novel approach to predicting product success in the technology market, with a specific focus on smartphones and laptops. Leveraging machine learning techniques and a meticulous data collection process, the study develops a predictive model that can serve as a tool for manufacturers, marketers, and decision-makers in the technology industry. The research employs a threshold-based feature analysis methodology to classify product success. This approach integrates domain knowledge and customer preferences, resulting in clear andinterpretable criteria for success. Feature engineering's importance is underscored, particularly in the creation of 'average rating' and 'percentage success' metrics.

The data collection process, encompassing product specification selection, is meticulously detailed. The robustness of this process ensures a comprehensive and relevant dataset for both smartphones and laptops, forming a solid foundation for the subsequent predictive modeling. The development, training, and evaluation of Random Forest Classifier models form the heart of the thesis. Their performance on test data is thoroughly analyzed and visualized , examining metrics such as accuracy, precision, recall, F1 score, and ROC AUC. The implications of these results are discussed in depth, offering insights into the predictive power and limitations of the models.

The versatility of the threshold-based feature analysis method is demonstrated through its application to laptops and phones. The process of adapting the method to different product categories is discussed, highlighting its broad applicability. A key feature of the study is the deployment of the predictive models using Streamlit, an open-source app framework. This interactive web application allows users to input specific product features and receive a prediction for product success, making the outcomes of this research practically applicable and user-friendly.

Inconclusion, this thesis contributes a valuable predictive model for product success

in the technology market. It offers a comprehensive exploration of the process, results, and implications of using machine learning techniques for predicting product success, and its deployment in a user-friendly web application. The study opens avenues for future research in this area, potentially extending the methodology to other product categories and incorporating more complex machine learning techniques.

Index Terms: *predicting, product success, technology market, smartphones, laptops, machine learning techniques, data collection process, predictive model, threshold-based feature analysis, user-friendly web applications.*

Introduction：

The way we depend on devices like smartphones, computers, and smartwatches for carrying out regular tasks is growing as technology rapidly transforms our world. These things, which were once thought to be only for fun, are now crucial elements in our daily companions that help us stay connected, organized, and entertained. They assist us in waking up, guide our way, and also keep an eye on our well-being It's tough to picture a day without them.

The manufacturers of these devices are always competing because they have gotten deeply embedded in our daily lives. Making the appropriate device is more important than simply making a device. The one that everyone wants their hands on, from the busy professional to the college student. This explains why the tech industry is so dynamic. Companies frequently search for new and exciting thoughts in an effort to compete with other companies with fresh and imaginative ideas.

1.1 Structure of the Thesis

The thesis' remaining sections are organized as follows: A review of the relevant literature is presented in Chapter 2, the research methodology is described in Chapter 3, the analysis of the findings is presented in Chapter 4, the findings are discussed in Chapter 5 in relation to the research objectives, and the thesis is concluded and potential research directions are suggested.

This revised introduction provides a clear, logical structure, beginning with the context and importance of the research, then moving on to the problem statement, research aims and objectives, research output ,and significance, and finally the structure of the thesis. Simply stating what will happen in the following chapters effectively sets up the remaining part of the thesis.

1.2 Contribution to the Academic Field

The findings of this study could make a meaningful academic contribution in the field of predicting the success of technology products. By developing an effective prediction model that includes both product features and user reviews, our research may significantly fill up a gap in the body of existing literature. The findings could provide a foundation for future research in this area, stimulating further academic debate and inquiry into predicting tech product success.

1.3 Practical Implications for the Tech Industry

The predictive model developed through this research could offer valuable insights for various stakeholders in the tech industry. The model can be put to use by-product creators to help them decide which features are expected to contribute to a product's success. The model could be used by marketers to forecast how customers react to various features of a product, enabling more specialized and successful marketing campaigns. The model may be used by administrators at digital businesses when making important choices such as which products to make investments in or the best way to price new products.

NEED OF THE STUDY

The use of machine learning in the tech industry has transformed the way devices such as smartphones and laptops operate and interact with users. Its diverse applications extend to several areas including predictive analytics, data mining and image processing(C. Wang et al., 2021).Machine learning plays a pivotal role in improving key features of smartphones. It is employed to optimize voice recognition, enhance camera quality, and improve battery life. By analyzing user behavior and system performance, machine learning algorithms can make real- time adjustments to these features, offering users an improved and more personalized experience(R et al., 2021).Similarly, laptops also leverage machine learning to enhance their performance and security. Machine learning algorithms can analyze user behavior to adapt system settings, thereby enhancing user experience and device security. This ability to adapt based on user behavior allows laptops to offer a personalized and optimized user experience(R et al., 2021).The potential of machine learning extends beyond basic device functionality. It is being utilized in the development of advanced technologies such as smart assistants, virtual *reality, and augmented reality. These technologies, which have found profound applications in smartphones and laptops, leverage machine learning to offer users immersive and intelligent experiences(Argade et al., 2021).*

In essence, the integration of machine learning in the tech industry has significantly improved the functionality and performance of devices like smartphones and laptops. It allows for continuous enhancements, providing users with more advanced experiences and capabilities, and shaping the future of the tech industry(Shamsuzzoha et al., 2022) (Sarker, 2021).

Research Methodology

3.1Introduction

This chapter explains how we conducted our study to predict product success in the technology market, focusing on smartphones and laptops. It covers data collection, preprocessing, and analysis steps, including feature analysis for smartphone and Laptops success. We discuss collecting product data, calculating success metrics, and training predictive models. We also explore model deployment using Streamlit. Each part offers clear explanations and codeexamples to understand our research process thoroughly.

3.2Data Collection

A technique known as web scraping is used for gathering data from websites. In the context of e-commerce, website scraping can be used to gather product data from many different websites (Ambre et al., 2019). One popular online scraping tool is BeautifulSoup, a Python package that makes it simple to parse and extract data from HTML and XML files(Ullah et al., 2018). It offers a straightforward and flexible API for extracting data from many formats, like JSON, XML, CSV, and others(Rahman & Tomar, 2021). The process of collecting complete product details from e- commerce websites and saving them in a database for analysis, later on, is able to be done by developers(Gulik et al., 2021) by using BeautifulSoup. This makes it simple for customers to compare products and costs among multiple websites, which could save both time and effort(Prakash & Rashid, 2017).

As part of the approach for gathering data for this study, a Python script was written to collect information from the Amazon online store using web scraping methods. We picked this platform for our study because it has lots of different products, well-organized information about them, and reviews from users. This makes it a trustworthy place to get data for our research.

3.2.2 Data Collection for Laptops

The data gathering for laptops using a Python script that uses web scraping methods to retrieve information from the Amazon platform, just like the data collection for mobile phones. However, this data collection specifically targetedcertain characteristics and features of computers.

To avoid being blocked by Amazon’s server, the script begins by setting request headers to resemble a real browser. The product information is then extracted from the parsed HTML content using helper methods that are defined to retrievethe webpage’s HTML code.

3.3Data Preprocessing

A key step in our research methodology is data preparation, which is the transformation of unstructured data into understandable form. Online scraping might generate raw data that is unpredictable inconsistent, or missing crucial information. As a result of data preparation, the raw data is cleaned, structured, and organized, so it's ready for the following steps of the data analysis pipeline.

The data preprocessing Python script uses a number of methods, such as handling missing values, data transformation, and data encoding. These procedures guarantee the accuracy and value ofthe data for later analysis.

3.3.1 Handling Missing Values for Smartphones

Due to a variety of factors, such as the absence of specific product details on the Amazon website, the raw data obtained by web scraping frequently contains missing numbers. Because they can produce biased or inaccurate results during data analysis, missing values must be handled carefully.

Depending on the situation, this code handles missing values differently. For the 'battery_type' column, for instance, if a value is absent, the code uses the 'fillna_by_manufacturer' function to fill it with the most typical battery type for that manufacturer. When it is unable to identify a common battery type, the rows with blank values for "battery_type" and "form_factor" are removed from the dataset.

3.3.2 Data Transformation for Smartphones:

Data transformation entails transforming data from its unprocessed form into one that is better suited for analysis. Among the activities that fall under this category are those that normalize numerical data, bin continuous data into discrete intervals, or extract relevant information fromtext data.

3.3.3 Data Headers

The following are the columns collected before data preprocessing for smartphones:

"product_id","mrp", "model_name","no of 4 stars" ,"no of 3 stars", "no of 2 stars," "no of 1 stars," "os," "ram," "inbuilt_storage," "dimensions," "weight," "battery_power," "battery_type," "camera," "warranty," "form_factor," and "manufacturer."

The following are the columns collected after data preprocessing for smartphones:

product_id,mrp,no of 5 star,no of 4 star,no of 3 star,no of 2 s tar,no of 1 star,ram,inbuilt_storage,weight,battery_power,battery_type,form_factor,length,width,height,os _n ame,os_version,phone_warranty(months),camera_count,cam_has_AI,cam_has_OIS,cam_has_ Zo om,cam_has_HDR,cam_has_Macro,cam_has_Portrait,main_camera_MP,is_success,no_ratings,a vg_rating.

cam_has_Portrait	main_camera_MP	is_success	no_ratings	avg_rating
0	50	TRUE	28467	4.110022131
0	12	TRUE	22473	4.090152628
0	13	TRUE	5049	3.888888889
0	12	TRUE	1293	4.173240526

Table 1-Column name headers for smartphones data.

The following are the columns collected before data preprocessing for laptops:

"product_id","mrp","model_name","screen_size,"display_resolution","os","hard_disk_typ e","h ard_drive_size","ram_memory","processor_brand","processor_name","processor_speed"," proce ssor_count","display_type","product_dimensions","batteries","form_factor","audio_details ","speaker_details","connector_types","graphics chipset","graphics type","graphics ram type","graphics details","brand","no of 5 star","no of 4 star","no of 3 star","no of 2star","no of 1 star":


```
{
    "product_id": "B0BV74GVWT",
    "mrp": 48990.0,
    "model_name": "E14 Gen4",
    "screen_size": "14 Inches",
    "display_resolution": "1080p",
    "os": "Windows 11 Home",
    "hard_disk_type": "SSD",
    "hard_drive_size": "512 GB",
    "ram_memory": "8 GB",
    "processor_brand": "Intel",
    "processor_name": "Core i3",
    "processor_speed": "1.2 GHz",
    "processor_count": "1",
    "display_type": "LED",
    "product_dimensions": "22.1 x 32.4 x 1.9 cm; 1.59 Kilograms",
    "batteries": "1 Lithium Polymer batteries required. (included)",
    "form_factor": "Laptop",
    "audio_details": "Headphones, Speakers",
    "speaker_details": null,
    "connector_types": "No Optical Drive",
    "graphics chipset": null,
    "graphics type": "Integrated",
    "graphics ram type": "Shared",
    "graphics details": null,
    "brand": "Lenovo",
    "no of 5 star": 9,
    "no of 4 star": 5,
    "no of 3 star": 0,
    "no of 2 star": 2,
    "no of 1 star": 3
},
```


Figure 2 - Column name headers for laptops sample data(JSON).

The following are the columns collected after data preprocessing for laptops:

product_id,mrp,os,hard_disk_type,ram_memory,processor_brand,processor_count,display_type,form_factor,screen_res_w,screen_res_h,length,width,height,weight,hard_drive_size_value, hard_drive_size_unit,battery_type,is_success

product_id	mrp	os	hard_disk_type	ram_memory	processor_brand	processor_count	display_type	form_factor	screen_res_w	screen_res_h	length	width	height	weight	hard_drive_size_value	hard_drive_size_unit	battery_type	is_success
B0BXPWS41N	70990	5	2	16	0	8	3	9	2880	1800	35.7	22.8	2	1.7	512	0	2	FALSE
B099ZZX3QW	53000	4	0	8	2	1	1	5	1920	1080	36.3	23.9	1.8	1.65	1	1	2	FALSE
B0BKT92YSL	144990	5	2	16	2	1	2	4	1920	1080	39.6	26	2.2	2.2	1	1	2	FALSE
B0BLH4C9Y6	33980	5	2	8	2	4	3	9	1366	768	32.5	21.6	2	1.55	256	0	2	FALSE

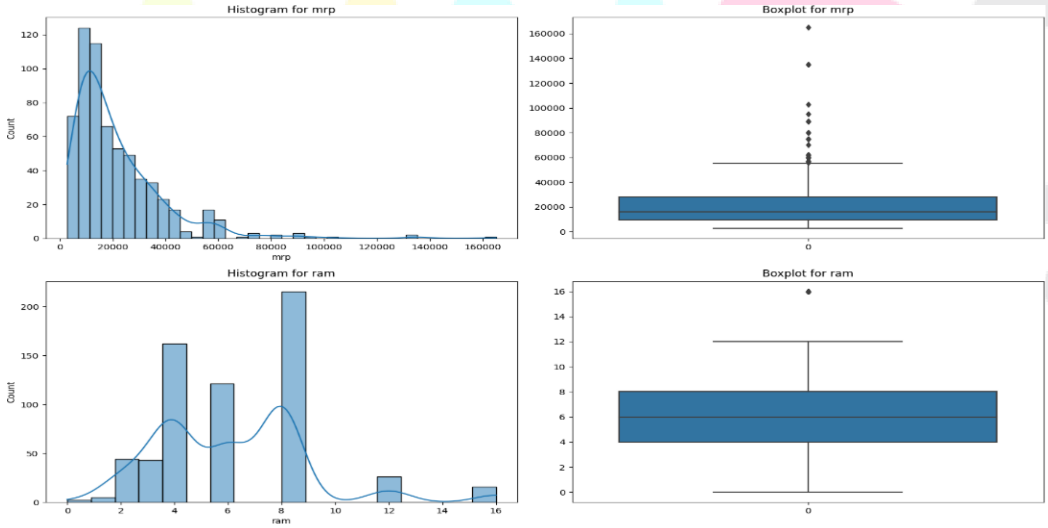
Table 2 - Column name headers for laptops sample data

3.3.4 Numerical Feature Analysis

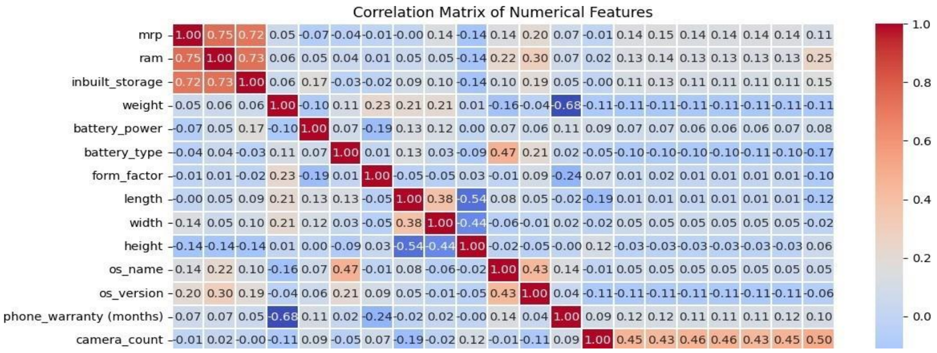
In this part of the exploratory data analysis, we analyze numerical features such as 'mrp', 'ram', 'inbuilt_storage', 'weight', 'battery_power', 'length', 'width', 'height', 'phone_warranty (months)', 'camera_count', and 'main_camera_MP'. This analysis includes creating histograms to understand the distribution of these features and boxplots to understand their spread and identify any outliers.

Histograms: Visual representation of data's value distribution using bars within defined ranges(bins).

Boxplots: Summarizes data's key statistics, displaying median as a line inside a box. Whiskers show data range, excluding outliers. Box spans middle 50% of data, while points outside whiskers may be outliers.



3.3.5 Correlation Analysis for phones:



Correlation reveals relationships between numbers. It shows how one changes as the other does. A coefficient of -1 to 1 represents strength. Close to 1 means strong positive, 0 weak, and near -1 strong negative correlation.

Descriptive statistics for laptops

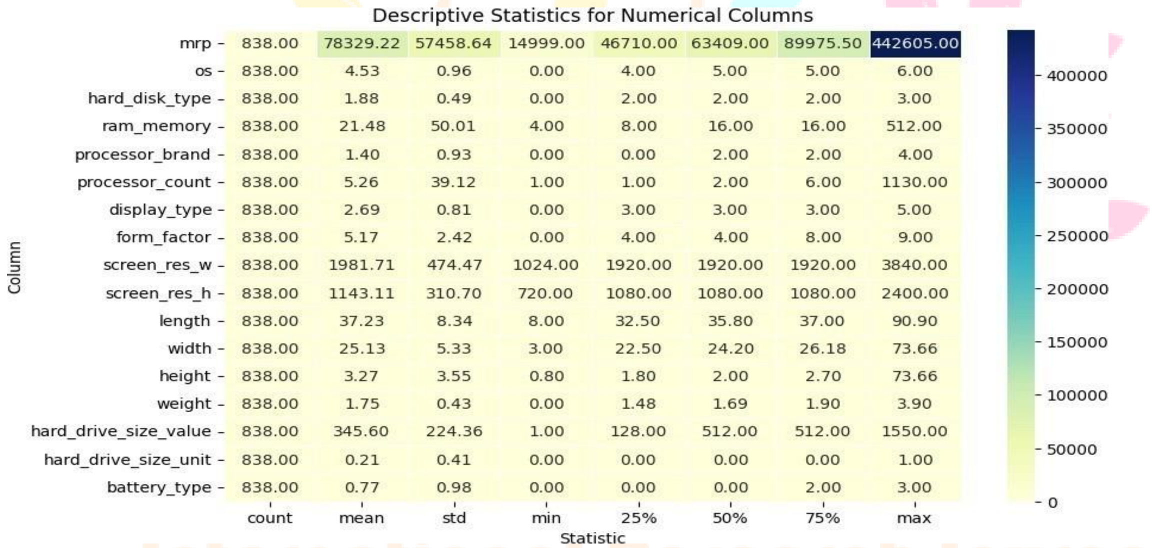


figure 8 Descriptive Statistics For Laptops

The table above shows the descriptive statistics for each numerical feature in the laptop dataset. Here is the explanation of the statistics:

Count: This shows the total count of available data in each column. All columns have 838 data points, indicating that there are no missing values.

Mean: The mean represents the typical value in each column. For example, the average price (mrp) for laptops is around 78329, and the average RAM capacity is about 21 GB.

Standard Deviation (Std): Std measures how much the data tends to spread out. A high standard deviation suggests data points are spread widely from the mean, while a low value means they'recloser to the mean.

Min: The minimum value in each column represents the smallest recorded data point. For example, the most expensive laptop costs 442605, and the laptop with the most RAM memory has 512 GB. These statistics provide a general overview of the numerical data in the dataset.

On this dataset, we will then conduct exploratory data analysis. Analyzing the distribution of the numerical and categorical variables, searching for outliers, and comprehending the connectionsbetween various parameters and the goal variable are all part of this process.

Conclusion

The literature review has critically analyzed and highlighted the key research areas related to predicting tech product success, specifically focusing on smartphones and laptops. It has illuminated the significance of feature-level rankings(Jerripothula et al., 2020), sentiment analysis(Kafi et al., 2019), threshold-based feature analysis(Malhotra & Sharma, 2021), and review-based rating prediction systems (Hasanzadeh et al., 2022)in understanding customer preferences and product performance. Additionally, it has emphasized the importance of machine learning as a transformative tool in the tech industry that can significantly enhance product prediction models(Shamsuzzoha et al., 2022).

However, the review also reveals notable research gaps. Despite the extensive studies on different aspects of tech products and the application of machine learning, a comprehensive and streamlined methodology that integrates threshold-based feature analysis with machine learning techniques to accurately predict tech product success remains largely unexplored. Also, a gap is seen in the lack of a user-friendly interface that connects users directly with a machine learning model for real- time predictions.(Heer & Kandel, 2012)

This research aims to bridge these gaps by proposing an innovative approach that not only integrates threshold-based feature analysis and machine learning techniques but also provides a user-friendly interface for real-time predictions. By doing so, this research can significantly advance the field of tech product success prediction (F.C.-W. Lo, 2000), contributing valuable insights for both consumers and industry stakeholders, and ultimately enhancing the customer decision-making process and overall satisfaction.

In conclusion, although the current research base is strong, there's plenty of room for more investigation and improvement in using machine learning and threshold-based feature analysis to predict the success of tech products. This study has the potential to greatly add to this area, presenting a fresh approach that's likely to improve both our knowledge and real-world use of these methods.

Future scope :

The things we couldn't explore in this study open doors for future research. To make this study better in the future, we could use information from various places and parts of the world to makethe model stronger and more useful for different situations.

Also, delving into other potential influencing factors like brand reputation, design, and customer reviews would enrich the dataset and possibly provide more accurate predictions. Applying natural language processing techniques to analyze reviews could offer valuable insights into thesubjective factors impacting a product's success.

Another promising direction for future work could be the application of deep learning algorithms. As our dataset grows, deep learning models might be better equipped to handle theincreasing complexity and provide improved accuracy in predicting product success.

In conclusion, this study has provided valuable insights into predicting product success based on product specifications and demonstrated the effectiveness of machine learning techniques in this domain. However, there are always opportunities for improvement, expansion, and refinement in future research endeavors.

References :

Allenby, G. M., Brazell, J. P., Howell, J. R., & Rossi, P. E. (2013). Economic Valuation of Product Features. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2338154

Ambre, A., Gaikwad, P., Pawar, K., & Patil, V. (2019). Web and Android Application for Comparison of E-Commerce Products. International Journal of Advanced Engineering, Management and Science, 5(4), 266–268. https://doi.org/10.22161/ijaems.5.4.5

Amin, H. H. (2021). Automated Adaptive Threshold-Based Feature Extraction and Learning for Spiking Neural Networks. IEEE Access, 9, 97366–97383. https://doi.org/10.1109/ACCESS.2021.3094262

Andi Purnomo, M. R., Azzam, A., & Uswatun Khasanah, A. (2020). Effective Marketing Strategy

Determination Based on Customers Clustering Using Machine Learning Technique. Journal of Physics: Conference Series, 1471(1), 012023. https://doi.org/10.1088/1742-6596/1471/1/012023

Argade, D. N., Pawar, S. D., Thitme, V. V., & Shelkar, A. D. (2021). Machine Learning: Review.

International Journal of Advanced Research in Science, Communication and Technology ,251–

256. https://doi.org/10.48175/IJARSCT-1719

Baccarella, C. V., Gerhard, D. A., & Voigt, K. I. (2010). Advertising for high-technology products in the product launch phase – a content-analysis. International Journal of Engineering Management and Economics, 1(4), 249. https://doi.org/10.1504/IJEME.2010.038645

Cristina, C., & Kurniawan, A. (2018). Sejarah, Penerapan, dan Analisis Resiko dari Neural Network: Sebuah Tinjauan Pustaka. Jurnal Informatika: Jurnal Pengembangan IT,3(2), 259–270. https://doi.org/10.30591/jpit.v3i2.890

Dadhich, A., & Thankachan, B. (2021). Opinion Classification of Product Reviews Using Naïve Bayes, Logistic Regression and Sentiwordnet

Hasanzadeh, S., Fakhrahmad, S. M., & Taheri, M. (2022). Review-Based Recommender Systems: A Proposed Rating Prediction Scheme Using Word Embedding Representation of Reviews. The Computer Journal, 65(2), 345–354. https://doi.org/10.1093/comjnl/bxaa044

Heer, J., & Kandel, S. (2012). Interactive analysis of big data. XRDS: Crossroads, The ACM Magazine for Students, 19(1), 50–54. https://doi.org/10.1145/2331042.2331058

Hosseini, Z., Mohammadi, S., & Safari, H. (2018). An Assessment of the Impact of Information

Technology on Marketing and Advertising. Engineering, Technology & Applied Science Research, 8(1), 2526–2531. https://doi.org/10.48084/etasr.1620

Hsee, C. K., Yang, Y., Gu, Y., & Chen, J. (2009). Specification Seeking: How Product Specifications Influence Consumer Preference. Journal of Consumer Research, 35(6), 952–966. https://doi.org/10.1086/593947