Using Data Mining Techniques to Analyze Student's Preferences for Smartphones:

A Comparative Study between iOS and Android Users in Higher Technical Institutes in Libya

Abdalkarim Erhab[1]

Higher Institute for Science and Technology, Algarabolli-Libya
karem_ar@yahoo.com

Hana Elammari[2]

Higher Institute for Science and Technology, Algarabolli-Libya
Hana.ali.bashir7@gmail.com

Khalleefah alzawam[3]

Higher Institute for Science and Technology, Algarabolli-Libya
khalleefah.alzawam@gmail.com

1.Abstract:

The intention of this particular study is to evaluate student smartphone preferences through data mining techniques on survey data collected from higher technical institute students in Libya. Special emphasis is placed on performing a comparative study with iOS and Android users to ascertain their purchasing decision drivers. An association rules mining approach using the FP-Growth algorithm was utilized to uncover user behavioral patterns and correlations among demographic features of users. The findings underscored the importance of the price, camera quality, security, and peer's advice influencing students' decisions, showing significant differences between the two user groups. This study has provided an analysis that can be used in marketing and product development planning focused to the students' needs.

Keywords : Data Mining, Smartphones, iOS and Android, Association Rules, FP-Growth, RapidMiner

2. Introduction:

As competition has increased in today's industrial era, data mining techniques to perform analysis of sales transaction data and identify regularity and patterns in consumer's behavior and preferences have become increasingly important for organizations to make insight based data-driven decisions [1].

In recent years there has been an observable increase in smartphone use by young people, making smartphones an important element of everyday life and, in many cases, a routine element in an educational context. The use of smartphones is increased due to their multi functionality. The smartphone functions might include social communication, entertainment and research and educational purposes [1]. As we begin to understand that young people's dependence on smartphones has increased, it is useful to be able to understand user behavior and user preferences when choosing their smartphones in order to identify factors that affect buying decision making or purchasing choices.

Data mining methods have shown, in particular association rule algorithms, to be a useful technique for discovering hidden patterns and relationships among the variables in a huge data set. They are used in a variety of fields including e-commerce and consumer-based behavior to uncover associations between items and identify frequent patterns in transaction data. Recent research has demonstrated the success of these techniques in mining the behavior of smartphone users, yielding extensive insights into their preferences and needs[2].

The smartphone market in Libya, as in many other places in the world, is growing and diversifying at an impressive rate, mirroring a variety of consumer preferences (predominately stated between two systems of operation; Android and iOS) are encountered. As students began to rely on smartphones for all types of learning, communication and entertainment, it is appropriate to explore the purchasing behavior factors that influence the student's decisions.

Data mining methods within this context provide the strength of analyzing such preferences and finding common behavioral patterns. In this study, among the available and familiar association rule algorithms, we went more specific in the FP-Growth, to uncover hidden and non-obvious relationships among variables in large data sets. The data sets were acquired from a structured questionnaire given to students attending higher technical institutes throughout Libya. The survey communal a range of determining influences students' preferences when engaging with smartphones. The FP-Growth model depends on statistically valid indicators such as support and confidence to represent the power and confidence of the relationships of concern, and thus provide a more devout understanding of the relationships between users' behavioral preferences and their technology-based choices.

3. Research Problem:

The research problem lies in the lack of utilization of data mining techniques in conducting analytical studies aimed at discovering behavioral patterns and relationships among the characteristics of smartphone users among students of higher technical institutes in Libya. This shortcoming negatively affects marketing decisions targeted at students, despite the widespread use of smartphones among them

4. Research Objectives:

This study aims to:

1. Collecting data through a questionnaire distributed to the students targeted by the study. This data includes the preferences selected by students when choosing smartphones.
2. Applying the FP-Growth algorithm to discover hidden and non-obvious relationships between variables in the large datasets obtained from the questionnaire directed at students enrolled in higher technical institutes in Libya.
3. Analyzing the obtained results and providing recommendations to application developers and device suppliers.

5. Research Questions: The research questions are as follows:

1. What are the main factors that influence students when choosing between iOS and Android operating systems?
2. Which algorithms can be used to identify common behavioral patterns among the study sample?
3. How accurate are the results obtained from applying data mining algorithms to classify students' preferences for iOS and Android operating systems?
4. How can the findings be utilized to support marketing strategies for smart device suppliers?

6. Research Scope :

This research targets students of the Higher Technical Institutes in Libya, where a sample of 180 students was selected for data analysis and result extraction.

7. Previous Studies :

There are numerous studies that focus on applying data mining techniques to analyze consumer behavior and preferences. Below is a brief overview of some of these studies:

7.1 (Finkelstein … et 2017) :

In this paper, It was used present a method to reveal various demographics and technical computer skills of smartphone users by their Internet traffic records, using machine learning classification models. We implement and evaluate the method on real life data of smartphone users and show that smartphone users can be classified by their gender, smoking habits, software programming experience, and other characteristics[3].

7.2 (Radzi. M …et 2021) :

The purpose of this study is to investigate the purchase decision of University Kuala Lumpur Business School students and its association with brand, price, features and social influence. The data set is collected through self-administered questionnaire. A total of 60 samples were recruited conveniently and all data collected were analyzed using SPSS software to generate the result. The findings revealed that brand, features and social influence variables were significantly influenced purchasing decision and price is found not giving any significant impact to the selection of smartphones among the university students. The finding will help marketers to understand the university consumer's Smartphone purchase decision as well as facilitate them in developing more attractive package to influence this segment which can be considered profitable for smartphones businesses and simultaneously could boost up the sales of smartphones [4].

7.3 (Parhusip. A 2020):

The study by Parhusip.A (2020) examined the use of the FP-Growth algorithm as a tool for analyzing smartphone sales data and identifying optimal promotional strategies. The study aimed to assist marketing managers in improving the efficiency of promotional campaigns by reducing costs and achieving sales goals more accurately. It focused on analyzing several variables related to user preferences, such as purpose of use, user experience, quality of service, and marketing analytics [5].

8. Data Mining:
Data mining is a process of selection, exploration and modeling of large amounts of data to find patterns or tendencies that are usually not aware of its existence. besides that data mining can also be interpreted as extracting new information taken from large chunks of data that help in decision making [6].

9. FP-Growth Algorithm :

The FP-Growth Algorithm, short for Frequent Pattern Growth, is an efficient data mining technique used to discover frequent patterns in large datasets. The method works by generating a compact data structure called the FP-tree (or frequent pattern tree), which stores the transactional information of the dataset. After creating the FP-tree the algorithm mines it to extract frequent patterns and, hence, has the advantage that it does not generate candidates itemsets. This makes the algorithm more scalable and fast compared to other algorithms such as Apriori. [7].
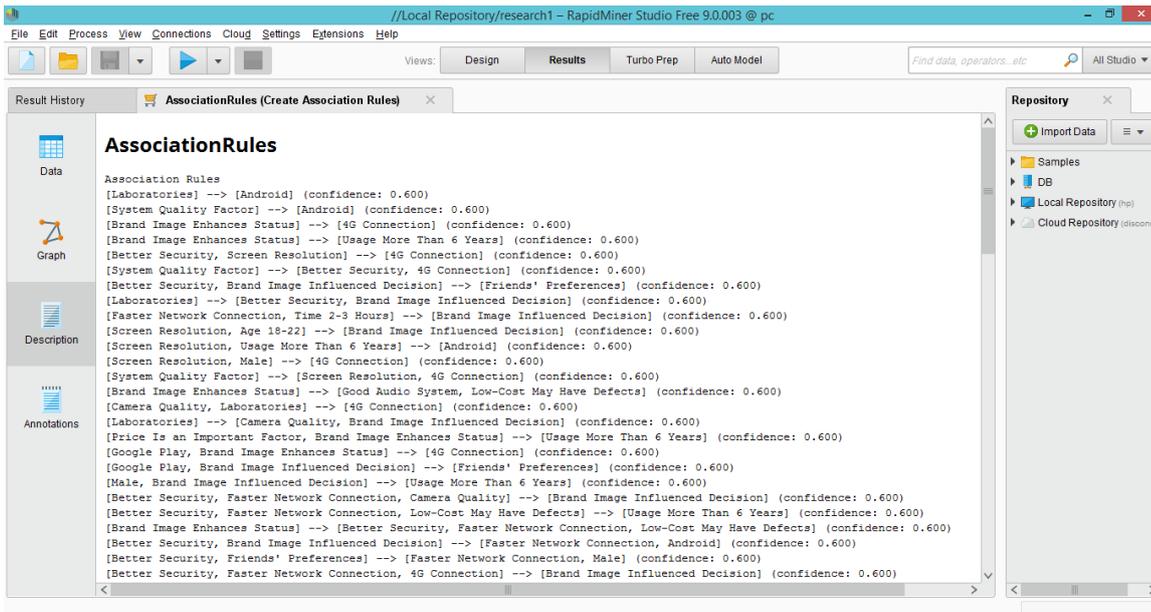
10. RapidMiner Platform :

RapidMiner is a leading, state-of-the-art software platform in the domain of data mining, offered by a company called RapidMiner. The platform provides users with an integrated environment that uses dozens of algorithms for data processing, machine learning (ML) and deep learning (DL) techniques, and predictive analytics. RapidMiner sees a lot of commercial applications and has been used in academic research to assist with scientific studies, education, training, rapid prototyping, and application development [8].

RapidMiner can help with any and all stage of the machine learning lifecycle, including data preparation, making model evaluations, analyzing model results, and tuning performance. A significant benefit to RapidMiner is that it can carry out 99% of your advanced analytical solutions using pre-built models which significantly reduces development time and errors by eliminating manual coding.

11. Implementation and Results

Data were collected through a questionnaire distributed to students of higher technical institutes in Libya. A total of 180 samples were gathered. The questionnaire was carefully designed to include a set of criteria relevant to the study, such as age, gender, purpose of phone usage, preferred features, current operating system, reasons for preference, influence of peers, and others.

The collected data was first entered into Microsoft Excel As shown in Figure(1):

(**Figure 1**) the Initial data before preprocessing

and subsequently imported into RapidMiner software for processing. The FP-Growth algorithm was employed to uncover hidden relationships among the variables within the dataset obtained from the questionnaire. The (Figure 2) explain the process of importing the Excel file into RapidMiner software , as well as the use of automated data preprocessing techniques prior to applying the FP-Growth algorithm to generate the desired results.

(**Figure 2**) the implementation of the FP-Growth algorithm

During the execution of the algorithm, more than 50 association rules were extracted from the dataset, as shown in (**Figure 3**). Each rule consists of four main components: Premises, Conclusion, Support, and Confidence.

(**Figure 3**) presents examples of the association rules that were discovered using the FP-Growth algorithm.

The next Figure (4) explain a set of statistical indicators used to assess the strength and reliability of the relationships extracted from the data:
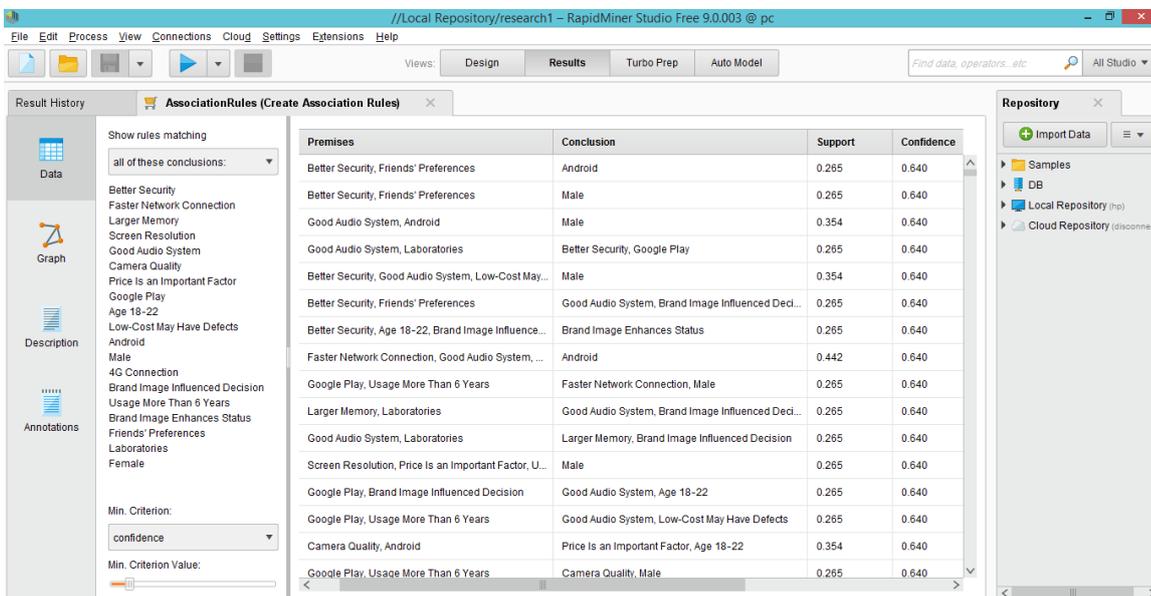
(Figure 4) explain a set of statistical indicators used to assess the strength and reliability of the relationships extracted from the data.

Definitions:

- **Premises:** A set of characteristics that lead to reaching a specific outcome.
- **Conclusion:** A single feature that appears when the premises are satisfied based on the characteristics.

- **Support:** The proportion of records that satisfy the rule out of the total number of records in the sample.
- **Confidence:** The probability of the conclusion occurring when the premises are true [9].

The support values ranged between 0.204 and 0.453, indicating a moderate frequency of these rules in the studied sample. The confidence level for most rules was 0.64, with a slight improvement in some rules reaching up to 0.685, which suggests a moderate strength of association.

Support is a measure used to express the prevalence of an item or a set of items within the database that is, the percentage of records that contain both the antecedent and the consequent of a rule.

Confidence is a measure used to estimate the reliability of an association rule that is, the percentage of records that contain the consequent Y among those that contain the antecedent X .

11.1 Examples of Extracted Rules from the FP- Growth Algorithm:

11.1.1 Example 1: The association between standards for technical specifications and longer-term usage data :

- *Premises*: "larger memory, camera quality, male"

- ***Conclusion***: "faster network connection, usage greater than six years"

- Support: 0.256

- Confidence: 0.649

- Analysis: The rule shows that it is male consumers who prefer high-quality cameras and larger storage capacity who also retain their choice of device for a longer duration and prefer speed of connection. This elevated usage of their devices confers a long-term investment behavior for technical devices and may reflect the commitment to a brand and a preference for high performance.

11.1.2 Example 2: How screen resolution and cost influenced female preferences:

- Premises: "Screen resolution, price as a key factor, female"

- Conclusion: "Laboratory use"

- Support: 0.215

- Confidence: 0.661

- Analysis: This rule shows female's interest in screen quality and cost associated with these device's value equations, which determine device categories or applications related to academic or practical settings.

11.1.3 Example 3: The influence of brand image and brand image on friend's preferences:

- Premises: "Screen resolution, price was an important issue, the brand image influenced the decision, the brand image enhanced the social status"
- Conclusion: "friend's preferences"
- Support: 0.2049
- Confidence: 0.6856
- Analysis: This rule demonstrates the effect of marketing and emotional attributes around brand image influences on user's behaviour with respect to trust in friend's recommendations. The relationship emphasizes the role of social proof and psychological factors has on making purchasing decisions.

11.1.4 Example 4 : Youth preferences with limited budgets

- Premises: "Increased security, camera quality, cost is a relevant factor, no cost with positive attributes"

- Conclusion: "Good noise system, 18-23 years old"

- Support: 0.315

- Confidence: 0.64

- Analysis: This rule indicates that youth willing to balance price and quality are often already accepting some technical shortcomings to gain key aspects of the product, i.e., security, and sound quality. This might support a focus on the construction of products for this consumer group.

11.1.5 Example 5: Observable purchasing behavior for male market segment:

- Premises: "Camera quality, price is an important factor, low price that may be defective, male"

- Conclusion: "Larger Memory, Android"

- Support: 0.227

- Confidence: 0.641

- Analysis: This rule shows male users are willing to make trades in performance versus cost where the male purchaser will accept inaccuracies in the device when the benefits of larger memory and Android platform have meaningful technical significance to them. The results further support the development of affordable, high performance devices for this demographic.

12. Conclusion:

The aim of this research was to examine students' preferences for purchasing smartphones utilizing data mining methodologies. In this study, FP-Growth Algorithm data mining methodology was used to assess data from students from higher technical institutes in Libya to explore relationships between the preferred operating system (Android, iOS), and a selection of personal and behavioral variables.

The findings indicated a strong correlation between Android use and a number of variables. These variables included ease of computer connection, price, applications, brand, male students, and the age range of between 18 and 22 years. The confidence for every rule extracted was 0.64, with a few rules slightly supporting a confidence level of 0.685. Moderate correlations were indicated with the initial rule set and additional variables such as length of use and prior user experience with computers.

The results also suggested that younger students, particularly males, also demonstrated a good understanding of the technical aspects of smartphones with a specific focus on android based operating systems that are perceived to have ease of use and compatibility with computers.

Finally, this research shaped a context for businesses that could consider how they can utilize this research. Data mining also provides business with a better understanding of consumer behavior when making purchase decisions in general and smartphone decisions, in particular, that can assist them in developing better methodology to select and create effective and targeted marketing strategies.

13. Recommendations:

This study suggests that smartphone manufacturers and marketers devise specific marketing campaigns aimed at smartphone-using segments, especially targeting young, male, tech-savvy customers. Also, Android users with background knowledge of computers and smartphones should be targeted with commercials stressing the price, availability of apps, and the ease of interlinking computers with smartphones.

Furthermore, it is recommended that other socio-geographic areas be considered which include additional measurements such as Lift and Gain in order to gain deeper insight into the preference patterns and the potency of the rules derived under various conditions .

14. References :

مراجع

[1] Prayugo. M. L & Wibowo. D. A & Hidajat. M. S & Mintorini. E.(2024) . *Data Mining Application for Analyzing Pattern of Customer Purchase Using Apriori Algorithm. Journal of Applied Intelligent System, pp.122–134.*

[2] Sarker.I.H & Salim. F. D(2018). Mining User Behavioral Rules from Smartphone Data Through Association Analysis**,** International Publishing AG, part of Springer Nature,   pp. 450–461.

[3] Finkelstein. A & Biton. R & Puzis. R & Shabtai. A (2017), *Classification of Smartphone Users Using Internet Traffic*. arXiv preprint arXiv:1701.0022, https://arxiv.org/abs/1701.00220 .

[4] Radzi. M & Isa.S.M & Rashid.R.A & Samah.A.A & Ramlan.N.H (2021) , *Factors Influencing the Selection of Smartphones Among University Students: An Insight from University Kuala Lumpur, Business School, Malaysia.* Asian Journal of Management, Entrepreneurship and Social Science.

[5] Parhusip. A (2020), *Penerapan Data Mining Dengan Algoritma FP-Growth Untuk Mendukung Strategi Penjualan Smartphone (Studi Kasus: PT. Oppo Indonesia).* Journal of Information System Research (JOSH), 209–215.

[6] S. Haryati, A & SudarsonHaryati, S& Sudarsono.A., & Suryana. E.(2015). Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5, Jurnal Media Infotama, 130–138, 2015

[7] *Activeloop: The FP-Growth Algorithm: A Scalable Method for Frequent Pattern Mining*, https://www.activeloop.ai/resources/glossary/fp-growth-algorithm .

[8] Hofmann.M & Klinkenberg.R(2013) , RapidMiner: Data Mining Use Cases and Business Analytics Applications, https://www.amazon.com/exec/obidos/ASIN/1482205491/acmorg-20.

[9] **GeeksforGeeks (2025).** *What is Support and Confidence in Data Mining,* https://www.geeksforgeeks.org/what-is-support-and-confidence-in-data-mining.

[10] Hailat.K & Alsmadi.S & Al-Abdi.Y & Galib,M (2024) .Assessment of Consumer Preferences for Mobile Phones Among University Students. *Journal of Lifestyle And SDGs Review, pp. 1-19 .*

[11] Celik.A.K & Eygu.H & Oktay.E (2015), A study on factors influencing young consumers' smartphone brand preference in Erzurum, Turkey , European Journal of Business and Economics , pp.24-31 .

[12] Ataş.A.H & Çelik.B(2019) , Smartphone Use of University Students: Patterns, Purposes, and Situations. *Malaysian Online Journal of Educational Technology,* pp.59-68 .

[13] Farheen.S, Consumer buying behaviour in smartphones(2017), *Reflections-Journal of Management,* pp.1–10 .

[14] Sarker. I. H & colman.A (2019), RecencyMiner: Mining recency-based personalized behavior from contextual smartphone data. *Journal of Big Data,* pp.1–20 .

[15] Al-Saggaf.Y & Rahman.M (2024), Using data mining to discover new patterns of social media and smartphone use and emotional states, *Social Network Analysis and Mining,* pp. 1–12 .

[16] Fitrianah.D & Zain.S.Y(2020), Analysis of consumer purchase patterns on handphone accessories sales using FP-Growth algorithm, *Proceedings of the International Conference on Engineering, Technology and Social Science (ICONETOS 2020)*, pp.149–158 .

[17] Lende. M & Rangga.A. A & Rato. K. W (2025), Application of the FP-Growth algorithm for mobile phone sales transactions, *JTH: Journal of Technology and Health,*pp. 132–140 .

[18] Putri.I.D & Yuhandri & Hardianto.R(2024), Application of the FP-Growth algorithm in consumer purchasing pattern analysis, *Journal of Computer Science and Information Technology,pp.* 44-49 .

[19] Kavitha.P.N (2022), Comparative analysis of Apriori and FP-Growth algorithms for frequent item sets , *International Journal of Advances in Management, Technology and Engineering Sciences,pp.* 9–15 .

[20] Hartono & Fitrianah.D(2020), "A Data Mining Approach For Product Promotion and Inventory Solution using FP-Growth Algorithm, International Journal of Computer Applications,pp.37–44.