

PREDICTION OF ANDROID HANDPHONE SALES DURING PANDEMIC USING NAÏVE BAYES AND K-NN METHODS BASED ON PARTICLE SWARM OPTIMIZATION

Endang Sri Palupi

Sistem Informasi
Universitas Bina Sarana Informatika
endang.epl@bsi.ac.id

Abstrak

Pada masa pandemic sebagian besar sekolah, kampus, dan tempat pendidikan melakukan kegiatan belajar mengajar secara online. Kegiatan belajar mengajar banyak dilakukan dengan menggunakan aplikasi zoom, google, webex, atau microsoft teams. Semua itu bisa dilakukan melalui laptop, bisa juga menggunakan Handphone (HP) sehingga kebutuhan akan laptop dan HP meningkat baik barang baru maupun barang yang bekas. Walaupun dimasa pandemic keadaan ekonomi menurun, banyak perusahaan yang mengalami kerugian, sehingga terjadi pengurangan karyawan dan menimbulkan angka pengangguran yang tinggi, kebutuhan akan HP android tetap tinggi. Selain untuk sarana pembelajaran jarak jauh secara online, HP android juga bisa digunakan untuk penjualan online melalui e-commerce, market place, media sosial, dan platform digital lainnya. Saat ini HP android mempunyai banyak pilihan dan sesuai dana yang kita miliki, dengan berbagai brand dan spesifikasi. Banyak merk mengeluarkan produk HP android dengan spesifikasi yang cukup bagus dan harga yang terjangkau, sehingga walaupun daya beli menurun akibat pandemi, penjualan HP android tetap banyak. Dalam penelitian ini penulis memprediksi penjualan HP android yang terbanyak menggunakan metode Naïve Bayes dan metode K-Nearest Neighbor berbasis Particle Swarm Optimization. Hasil dari penelitian ini menggunakan Algoritma Naïve Bayes yaitu sebesar 74.92%, sedangkan menggunakan Algoritma K-Nearest Neighbor berbasis Particle Swarm Optimization nilai akurasi sebesar 81.33%.

Kata kunci: Android, K-Nearest Neighbor, Naïve Bayes

Abstract

During the pandemic, most schools, campuses, and places of education conducted online teaching and learning activities. Many teaching and learning activities are carried out using the Zoom, Google, WebEx, or Microsoft Teams applications. All of that can be done through a laptop, you can also use a cellphone (HP) so that the need for laptops and cellphones increases, both new and used goods. Even though during the pandemic the economic situation was declining, many companies suffered losses, resulting in a reduction in employees and causing a high unemployment rate, the need for Android phones remains high. In addition to online distance learning facilities, Android phones can also be used for online sales through e-commerce, market places, social media, and other digital ceilings. Currently, Android phones have many choices and according to the funds we have, with various brands and specifications. Many brands issue android cellphone products with pretty good specifications and affordable prices, so that even though purchasing power has decreased due to the pandemic, sales of android cellphones are still high. In this study, the author predicts the highest sales of android cellphones using the Naïve Bayes method and the K-Nearest Neighbor method based on Particle Swarm Optimization. accuracy of 81.33%.

Keywords: Android, K-Nearest Neighbor, Naïve Bayes

INTRODUCTION

When the COVID-19 pandemic hit Indonesia, sales of Android cellphones actually increased, both new and used cellphones. The need for an Android cellphone is related to online

teaching and learning activities from kindergarten to university level students, online business that many people do during a pandemic, for content creators and various activities can be done using an Android cellphone during this pandemic. The author conducted a study to predict the most sales



of Android cellphones during the covid19 pandemic to find out the biggest target market for Indonesian people is the lower middle class. Android cellphones at affordable prices but with good specifications and performance are the most widely purchased flagship by the public. Various brands and types of Android phones with competitive quality and prices are currently very busy on the market, so people have many choices depending on their needs and available funds. No matter how good the sophistication of a smartphone, it will be redundant and useless if it doesn't fit your needs. (Solihin, 2017)

Prediction is an attempt to predict future conditions through testing past conditions. Forecasting sales means determining the estimated amount of sales volume, even determining to make decisions or policies in accordance with the results of the sales predictions, before this scientific research is carried out the potential sales and market area controlled in the future. (Wibowo, 2018). With this research, it is possible to predict sales of HP that are selling well and not selling well according to the interests of buyers, so that sellers can prepare inventory of goods in the future according to the interests of buyers and are expected to further increase sales and reduce the risk of loss due to stock items that are not selling well.

In a previous study, Eka Pandu Cyntia and Edi Ismanto in 2018 conducted a study entitled "C4.5 Decision Tree Algorithm Method in Classifying Sales Data for Fast Food Outlets" (Cynthia & Ismanto, 2018). The purpose of this study is to classify fast food outlets that are popular (selling) and less popular (less selling) using Algoritma C4.5. The results of this study indicate Price - Amount Sold - Food Menu (Rice Bento = Less Selling, Dada = Selling) with the weight of each attribute: Price (0.738), Menu Type (0.067), Amount Sold (0.156), Status Sales (0.040). In this study, Eka Pandu and colleagues only used one C4.5 algorithm and displayed only the decision tree, not calculating its accuracy. The results of this study are the value of each attribute used and what foods are selling well and not selling well. While in this study the authors predict sales of Android phones using 2 methods of the Decision Tree algorithm and the K-NN algorithm as a comparison and calculate the accuracy of each algorithm.

Another study was written by Juna Eska in 2016 from STMIK Royal Ksieran entitled "Application of Data Mining for Wallpaper Sales Prediction Using the C4.5 Algorithm" (Eska, 2016). The result of this research is that the highest factor influencing sales is the number of wallpaper motifs. The factors of Price, Size, Quality of Materials, and

Colors do not affect purchases because wallpapers with high prices, small sizes, good quality materials, and few colors are still in demand by customers. This study only uses the C4.5 algorithm method, previously calculated the Gain and Entropy values manually using excel, then for the decision tree using Rapidminer, the results of the study are that the highest factor affecting wallpaper sales is motifs. While the writer uses the C4.5 algorithm, the K-NN algorithm as a comparison, the results of this study are the accuracy values of each algorithm and the conclusion is that the PSO-based K-NN algorithm has a greater accuracy value than the C4.5 algorithm.

The research entitled Prediction of Honda's Best Selling Products With Classification Method Using the C4.5 Algorithm (Case Study: Sales Data of PT Prospect Motor, Cikarang) was written by Aswan S Sunge and Heri Fidiawan in 2019. (Sunge & Fidiawan, 2019) From the research The accuracy obtained is 67.5%, while in this study the author uses 2 algorithms as comparison material, namely the C4.5 algorithm and the PSO-based K-NN algorithm with greater accuracy using the PSO-based K-NN algorithm with an accuracy value of 81.33%.

In 2020 Alfian Faiz I and Sulastri conducted a research entitled "Classification of Android Application Sales Using the C4.5 Algorithm". This study classifies sales of applications that can be used on Android. Three classification experiments resulted in different accuracy values. The highest accuracy value is obtained from experiments with 210 training data and 90 testing data, which shows an accuracy rate of 73.3%. Rating is the attribute that most influences the application that is classified as selling or not selling. The difference is that the research classifies sales of android applications, while in this study the authors predict sales of cellphones using android applications. This research only uses the C4.5 algorithm method, there is no comparison with other algorithms (Alfian Faiz Izzulhaq1, 2020)

In 2020 Ahmad Zakir and colleagues conducted a study entitled "Application of Data Mining for Classification of Best Selling Food Sales Data with Algorithm With C4.5 Algorithm". The research was conducted at a burger shop, and implemented using the C4.5 Algorithm with PHP programming language and MySQL database. The result of this research is to produce a system that can determine which foods are selling well and not selling well using the C4.5 algorithm. The difference with this study is that the author uses two algorithms as a comparison with the results of the greatest accuracy value using the PSO-based K-NN algorithm of 81.33%, and the author uses the

RapidMiner framework in the implementation of the algorithm. (Zakir, Ndruru, & Hadinata, 2020)

In 2020 Ismasari Nawangsih conducted a study entitled "Application of the Naïve Bayes Algorithm to Determine the Classification of the Best Selling Products in Credit Sales" in this study the authors performed calculations manually using the Naïve Bayes formula then calculated using Rapid Miner. The result of this research is that the name of the Telkomsel Pulsa product is the best-selling product and the accuracy value is 97.50% so that Naïve Bayes is a fairly good method in classification. While the authors in this study directly calculate using Rapidminer to calculate accuracy using 2 algorithm methods, namely the Naïve Bayes algorithm and the PSO-based K-NN algorithm, the results using the PSO-based K-NN algorithm have a greater accuracy value of 81.33%. (Ismasari Nawangsih1), 2020)

This study aims to predict the most sales of Android phones during the pandemic. In this pandemic period, the need for Android cellphones is increasing, and with this research, it can be used as a reference for buying cellphones or buying and selling cellphones, what Android cellphones are the most popular and currently selling the most. In addition, from the HP manufacturer's point of view, this research can also see what needs are most needed by the community in using HP in terms of features, price, specifications, and quality. Predicting market needs is very difficult which is a problem faced by distribution companies. To find market needs, it is necessary to identify customer characteristics. (Faradillah, 2013)

RESEARCH METHODS

This research uses quantitative techniques. According to Sugiyono, quantitative research methods can be interpreted as research methods based on the philosophy of positivism, used to examine certain populations or samples. (Sugiyono, 2016). Quantitative research can be defined as a process of finding knowledge by using data in the form of numbers as a tool to analyze information about what you want to know. This research method translates data into numbers to analyze the findings. The variables used in classifying best-selling products are types of goods, brands and prices with the target class being in demand and not selling well. These three variables are the benchmarks used in researching best-selling products in trading companies. (Asmaul Husnah Nasrullah, 2021).

Types of research

This type of research is quantitative, by taking data on sales of Android cellphones during the period of the pandemic in Indonesia. Where during the pandemic the need for Android cellphones actually increased for the teaching and learning process, business processes such as creating content or selling online through social media, marketplaces, and e-commerce.

Time and Place of Research

This research was conducted in June 2020 when the pandemic occurred in Indonesia, using HP sales data for the period June 2020 to May 2021 at ITC Cempaka Mas, ITC Roxy Mas, and several e-commerce sites in Indonesia.

Research Target / Subject

The target of this research is the productive age of Android HP users from 12 years to 50 years. Researchers conducted direct interviews, observations, and distributed google forms on whatsapp groups of students, junior high and high school students regarding the needs of HP and cellphones they currently have or want/have purchased.

Procedure

This framework represents the steps and procedures that will be carried out in this research process.

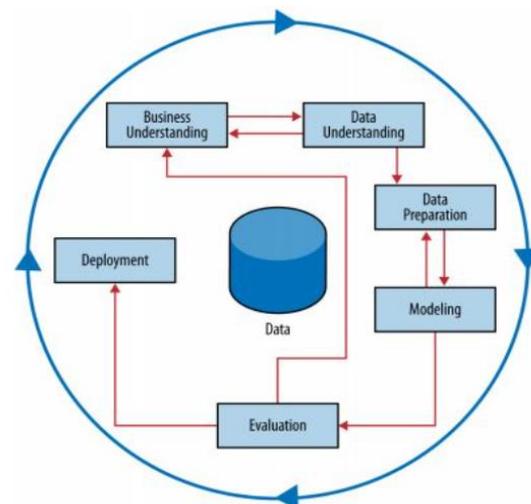


Figure 1. Cross-Industry Standard Process for Data Mining (CRISP-DM)

1. Business Understanding

This study aims to predict the sales of android cellphones during the pandemic, which cellphones are the best and the least sold. It is hoped that with this research distributors can understand the needs of the community so that

in the future they can stock goods and sell goods according to the needs and desires of the community, thereby increasing sales and avoiding losses due to selling cellphones that are not selling well.

2. Data Understanding

At this stage, the authors take HP sales data at HP stores in the ITC Cempaka Mas and ITC Roxy areas. conduct observations and interviews with HP sellers and buyers. As well as making observations on e-commerce sales of android cellphones, what android cellphones sell best. Sales data was taken when the pandemic occurred, which is between June 2020 to May 2021.

3. Data Preparation

At this stage, all the data that has been collected there are some useless attributes that must be removed using Delete Useless Attributes. The attributes used after this process are Brand, Type, Price, RAM, Memory.

4. Modeling

The author does the modeling using Rapid Miner Studio using the Naïve Bayes algorithm and the PSO-based K-NN algorithm to get the accuracy value. The author uses 2 algorithms as a comparison to get the best accuracy value.

5. Evaluation

From the modeling results, it can be seen that using the PSO-based K-NN algorithm modeling to get the best accuracy value compared to using the Naïve Bayes algorithm.

6. Deployment

The best results from this modeling use the PSO-based K-NN algorithm with an accuracy value of 81.33%.

Data, Instruments, and Data Collection Techniques

Data collection using the following techniques:

1. Interview

Interviews are used as a data collection technique to find problems that must be investigated and also if researchers want to know things from respondents more deeply. about behavior, and the meaning of that behavior (Sugiyono, 2016). The author visited several HP shops at ITC Roxy Mas and ITC Cempaka Mas, to get data on any HP sales in the period June 2020 to May 2021. Where in June 2020 and May 2021 there was an Eid moment so that sales increased. The author also interviewed shoppers to find out what cellphones they wanted to buy or which had been purchased.

2. Observation

Observation is a data collection method that uses direct or indirect observation (Yatim Riyanto, 2010). The author observes the sale of android cellphones in e-commerce such as shopee, tokopedia, JD.id, bhinneka.com, and others. By checking the number of items sold and the rating given by the buyer, as well as checking the best seller type of HP at that time in e-commerce.

3. Questionnaire

A questionnaire is a written list of questions given to the subject under study to collect the information needed by the researcher (Kusumah, 2011). The author shares a google form link that contains questions related to the needs of HP and HP currently owned or that want / have been purchased on the student WhatsApp group, high school and junior high school student whatsapp group. The result is that 110 people filled out the google form.

Data analysis technique

Put forward quantitative research, namely a research approach that uses a lot of numbers, starting from collecting data, interpreting the data obtained, and presenting the results. (Arikunto, 2006). In this study, the author uses quantitative data analysis techniques, namely interviews with cellphone sellers and cellphone shop visitors, observing cellphone sales to various e-commerce, and distributing google form questionnaires to students, middle and high school students regarding the needs of cellphones that they already have or want to buy. A total of 110 people filled out the google form and the total data obtained was 500 datasets.

With this research, the author wants to classify the types of android cellphones that are sold the most during the pandemic, where during the pandemic the community's economy is also experiencing a decline, but the need for android cellphones is also increasing with the online teaching and learning process, and activities that many people do during the pandemic. i.e. like content creators or selling online.

RESULTS AND DISCUSSION

The following sales dataset used in this research is 500 data:

Table 1. Sales Data



DATA PENJUALAN HP ANDROID JUNI 2020 - MEI 2021						
N O	MERK	TYPE	HARGA (IDR)	RAM (GB)	MEMOR Y	BEL I
1	XIAOMI	Redmi 8A	1,400,000.00	4	256	NO
2	SAMSUN G	GALAXY A11	1,900,000.00	3	512	YES
3	SAMSUN G	GALAXY A21s	2,900,000.00	3	128	NO
4	XIAOMI	Redmi Note 8 Pro	10,200,000.00	4	256	YES
5	XIAOMI	Redmi Note 8	11,000,000.00	4	256	YES
6	SAMSUN G	A51	4,700,000.00	3	256	NO
7	XIAOMI	Redmi 9A	1,100,000.00	3	256	NO
8	SAMSUN G	GALAXY M11	1,400,000.00	3	256	NO
9	XIAOMI	Poco X3 NFC	2,800,000.00	4	512	YES
10	REALME	C20	1,200,000.00	2	128	YES
11	INFINIX	Hot 10s	1,700,000.00	4	128	YES
12	SAMSUN G	GALAXY A12	2,000,000.00	4	512	YES
13	Tecno	Spark 7 Pro	1,700,000.00	6	512	YES
14	XIAOMI	Redmi Note 10	2,300,000.00	4	128	YES
15	XIAOMI	Poco M3 Pro 5G	2,500,000.00	4	128	YES
16	XIAOMI	Note 10 Pro	2,500,000.00	8	128	YES
17	INFINIX	SAMSUN GALAXY A325G	3,800,000.00	8	512	NO
18	SAMSUN G	Poco X3 Pro	3,500,000.00	8	512	NO
19	XIAOMI	Pro	2,700,000.00	8	256	YES
20	OPPO	A54	2,500,000.00	4	256	YES
21	OPPO	A15	1,700,000.00	4	256	YES
22	XIAOMI	Redmi 9C	1,500,000.00	4	128	YES
23	SAMSUN G	GALAXY A02	1,800,000.00	2	512	NO
24	VIVO	Y12s	1,800,000.00	4	256	YES

	true BELI	true NO	true YES	class precision
pred. BELI	0	0	4	0.00%
pred. NO	3	348	359	49.01%
pred. YES	0	99	1043	91.33%
class recall	0.00%	77.85%	74.18%	

Figure 3. The results of the accuracy of the Naive Bayes Algorithm

Figure 3 shows the accuracy results using the Naive Bayes algorithm of 74.92%, with True NO 77.85% and True YES 74.18%.

KNN algorithm based on Particle Swarm Optimization

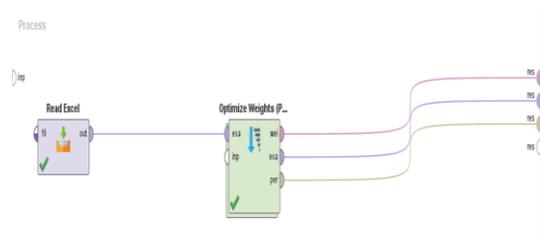


Figure 4. Design KNN algorithm based on Particle Swarm Optimization

Figure 4 is a design for weighting the KNN Algorithm based on Particle Swarm Optimization using 500 datasets.

Naïve Bayes classification

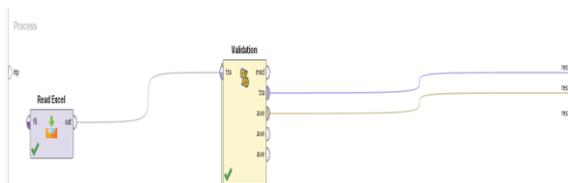


Figure 1. Naïve Bayes Algorithm

Figure 1 is modeling using the Naïve Bayes algorithm using 500 datasets.

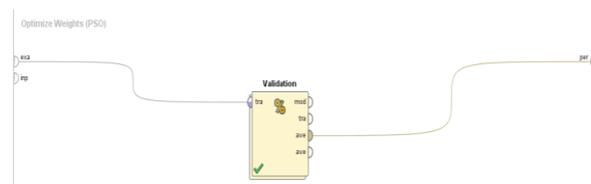


Figure 5. Validation of the Particle Swarm Optimization-based KNN Algorithm

In Figure 5 is the validation process of the PSO-based K-NN algorithm to increase the weight so that the accuracy value is better.

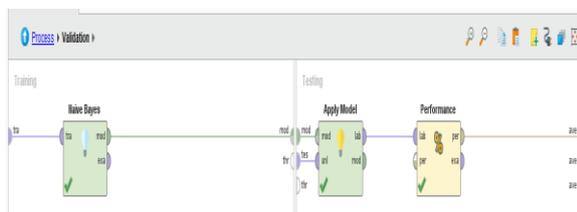


Figure 2. Nave Bayes Algorithm Validation Process

In Figure 2 is the validation process using the Naïve Bayes algorithm modeling to get the accuracy value.

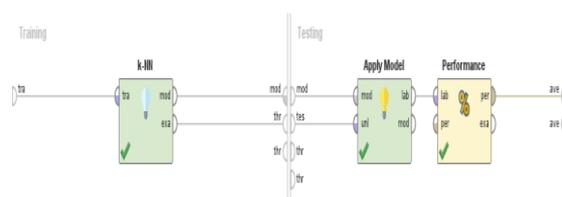
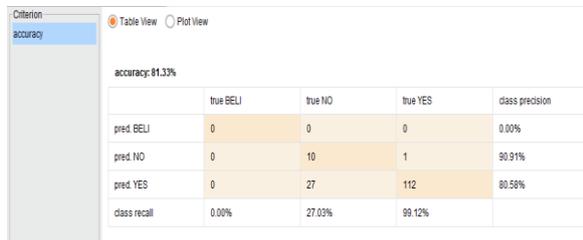


Figure 6. Design KNN Algorithm Based on Particle Swarm Optimization



After the weighting process using PSO, in Figure 6 the modeling using the PSO-based K-NN algorithm is ready to run to get a better accuracy value.



	true BELI	true NO	true YES	class precision
pred. BELI	0	0	0	0.00%
pred. NO	0	10	1	90.91%
pred. YES	0	27	112	80.58%
class recall	0.00%	27.03%	99.12%	

Figure 7. Accuracy results of the Particle Swarm Optimization-based KNN Algorithm

The accuracy results in Figure 7 using the PSO-based K-NN algorithm are better at 81.33% with True NO 17.03% and True YES 99,12%.

CONCLUSIONS AND SUGGESTIONS

Conclusion

The result of this study is the classification using the Naïve Bayes Algorithm is an accuracy value of 74.92% with a true NO recall class of 77.85% and true YES 74.18%. While the classification results using the Particle Swarm Optimization-based KNN Algorithm have an accuracy of 81.33% with a true NO 27.03% recall class and 99.12% true YES. Classification using the Particle Swarm Optimization-based KNN algorithm has higher results than using the Naïve Bayes Algorithm. Dengan demikian algoritma K-NN berbasis PSO bisa digunakan sebagai metode prediksi penjualan dengan hasil akurasi yang tinggi, sehingga dapat membantu meningkatkan penjualan.

Suggestion

Future research can use more varied data on questions about the applications needed, battery strength, camera needs and so on. Correspondence segments interviewed can be from business people who use Android cellphones, not just students and students.

REFERENCES

Alfian Faiz Izzulhaq1, S. (2020). Klasifikasi Penjualan Aplikasi Android. *Proceeding SENDIU*, (2019), 978–979.

- Arikunto, S. (2006). *Suatu Pendekatan Praktik (Revisi VI)*. Jakarta: PT Rineka Cipta.
- Asmaul Husnah Nasrullah. (2021). Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris. *Jurnal Ilmiah Ilmu Komputer*, 7(2), 45–51.
- Cynthia, E. P., & Ismanto, E. (2018). Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASI)*, (3) Juli(July), 1–13.
- Eska, J. (2016). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5. 2*. <https://doi.org/10.31227/osf.io/x6svc>
- Faradillah, S. (2013). Implementasi Data Mining Untuk Pengenalan Karakteristik Transaksi Customer Dengan Menggunakan Algoritma C4. 5. *Pelita Informatika Budi Darma*, 5(3), 1–5.
- Kusumah, W. D. D. (2011). *Penelitian Tindakan Kelas*. Jakarta: PT Indeks.
- Solihin, S. R. (2017). 10 Tips Panduan Memilih & Membeli HP Android Berkualitas. Retrieved from Septian website: <https://www.septian.web.id/10-tips-panduan-memilih-membeli-hp-android-bagus-berkualitas-html/>
- Sugiyono. (2016). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.
- Sunge, A., & Fidiawan, H. (2019). Prediksi Produk Laris Mobil Honda Dengan Metode Klasifikasi Menggunakan Algoritma C4. 5 (Studi Kasus: Data Penjualan Sales PT Prospect Motor, Cikarang). *Jurnal SIGMA*, 9(4), 97–103. Retrieved from <https://jurnal.pelitabangsa.ac.id/index.php/sigma/article/view/461>
- Wibowo, D. A. (2018). Prediksi Penjualan Obat Herbal Hp Pro Menggunakan Algoritma Neural Network. *Technologia: Jurnal Ilmiah*, 9(1), 33–41. <https://doi.org/10.31602/tji.v9i1.1100>
- Yatim Riyanto. (2010). *Metodologi Penelitian Pendidikan*. Surabaya: SIC.
- Zakir, A., Ndruru, Y., & Hadinata, E. (2020). Penerapan Data Mining Untuk Klasifikasi Data Penjualan Makanan Terlaris Dengan Algoritma C45. *Jurnal Ilmiah Teknologi Informasi Dan Robotika*, 2(2), 7–12. Retrieved from <http://jifti.upnjatim.ac.id/index.php/jifti/article/view/33>