

Application of the K-Nearest Neighbor (KNN) Algorithm in Machine Learning to Predict the Selection of Undergraduate Study Programs Based on New KIP Lecture Students

I Gusti Prahmana^{1*}, Adek Maulidya², Kristina A. Sitepu³, Razaq Habibi⁴

STMIK Kaputama

igustiprahmana4@gmail.com^{1*}, adek.maulidya@gmail.com², kannatasia88@gmail.com³, rozaqhabibi456@gmail.com⁴

Abstract

Higher education plays a vital role in shaping the future of individuals and society. Choosing the right study program is an important decision for every student, because it will affect their career path and personal development. The KIP Lecture program is present as a government initiative to provide higher education opportunities to students from underprivileged families. However, with the many options of study programs available, new students often have difficulty in determining the study program that suits their interests and abilities. On the other hand, the data of new students that is quite complete and varied opens up opportunities to use machine learning technology in helping the study program selection process. The K-Nearest Neighbor (KNN) algorithm as one of the simple and easy-to-implement machine learning algorithms has the potential to provide more accurate recommendations for the selection of study programs based on student data at STMIK Kaputama. Therefore, this study focuses on analyzing the use of the KNN algorithm in machine learning to predict the selection of undergraduate study programs. This research aims to identify existing problems, evaluate the effectiveness of KNN in this context, and provide solutions that can be implemented to improve the study program selection process for new students who receive KIP Lecture. It is hoped that it can provide recommendations for the selection of study programs that are more accurate and relevant for new students who receive KIP Lecture at STMIK Kaputama. In addition, this solution can also increase the effectiveness of academic guidance and assist students in achieving better academic and career success.

Keywords: Machine Learning, Application of K-Nearest Neighbor Algorithm, Prediction, Study Program, KIP Lecture

1. Introduction

In providing assistance, whether cash assistance or in-kind assistance, small or large, requires a consideration that must be researched first. Both physically and socially whether the prospective recipients are people who deserve to receive the smart Indonesia card, where the purpose of giving this smart card is intended for students who are really in need. With KIP, underprivileged students can continue their education up to the high school level. However, in granting KIP, the government makes conditions or criteria to be able to decide whether the prospective recipient of the smart Indonesia card (KIP) is eligible to receive KIP or not. The government sees that the opportunity to continue studying in excellent study programs at the best campuses throughout Indonesia must be given to all Indonesian children. PIP is assistance in the form of cash, expansion of access, and learning opportunities from the government given to students and students from families who cannot afford to finance education. In 2020, PIP, which was originally only given to children of primary and secondary education age, is now given to students under the name of the Indonesia Smart Lecture Card (KIP Kuliah) [1]. We hope that the nation's shoots that complete their studies will later be able to improve family welfare and contribute to building the country. The Ministry of Education and Culture, through the Center for Education Financing Services (PLPP/Puslapdik), has again opened opportunities for new students throughout Indonesia to become recipients of the 2024 KIP Merdeka Lecture until they graduate according to their level of study. Higher education plays a vital role in shaping the future of individuals and society. Choosing the right study program is an important decision for every student, because it will affect their career path and personal development. The KIP Lecture program is present as a government initiative to provide higher education opportunities to students from underprivileged families. However, with the many options of study programs available, new students often have difficulty in determining the study program that suits their interests and abilities. A strong commitment is needed from all parties so that the KIP Lecture program can run well, systematically, accountably, and on target to the nation's children who really need it according to the requirements. At the same time, it needs to be guided by strong regulations and a commitment to be implemented. Based on the above thinking, it is necessary to establish Guidelines for the Indonesia Smart Lecture Card Program in Higher Education with the aim of realizing the implementation of the KIP Lecture program that is orderly, transparent, accountable, and on target. As a commitment to the Indonesian government to educate the lives of Indonesian citizens. In this digital era,

machine learning technology has been widely used to assist in decision-making in various fields, including Education[10]. The K-Nearest Neighbor (KNN) algorithm is one of the simple but effective machine learning algorithms for making predictions based on existing data. In this context, the use of KNN is expected to help predict the selection of the most suitable study program for new students who receive KIP Lectures at STMIK Kaputama, by considering various factors such as academic background, interests, and career prospects. This study aims to analyze the use of the KNN algorithm in predicting the selection of undergraduate study programs. Through this research, it is hoped that important patterns and factors that affect the selection of study programs can be found, as well as provide more accurate recommendations to new students in choosing the appropriate study program. Thus, this research not only contributes to the development of science in the field of machine learning and education, but also provides practical benefits for students and educational institutions.

2. Page layout

2.1. Machine Learning

Developing Machine Learning (ML) and Artificial Intelligence (AI) systems that generate justice and equality is an important topic[8], especially since they are used in high-risk situations such as recruitment. The development of machine learning will follow the following approach:

1. Data Collection: Collects historical datasets that include code metrics, change history, and test results from previous software development projects.
2. Data preprocessing: Cleaning and preparing data for machine learning model training, including handling missing data, feature normalization, and advanced processing such as dimension reduction if needed.
3. Model Selection and Training: Select several machine-learning models such as Random Forest, SVM, and Neural Networks to be trained on preprocessed data. Model performance evaluation is carried out using cross-validation techniques to ensuring reliability and accuracy.
4. Model Validation and Testing: Test the model on a separate test dataset to validate the model's predictive and generalization capabilities against new data. Complementation and Integration: Integrating trained machine-learning models into existing software development pipelines or workflows, so that they can be used routinely in the development process.
5. Monitoring and Maintenance: Regularly monitor the model's performance and make adjustments as needed to ensure the model remains effective over time and changes in the software development environment.

By following these plans and designs based on the theory that has been explained, machine learning-based can improve the quality and efficiency of software development[6]. Machine Learning (ML) is a method, dataset, and optimization strategy. They can be accessed through standard channels defined by configuration files, Python dictionaries, or independently created. It provides a standard platform for experimental fairness testing[2].

2.2. Algoritma K-Nearest Neighbor (KNN)

Classification based on information systems is a technique of mapping (classifying) data into one or several predefined classes. There are many classification techniques that can be used, including K-Nearest Neighbor (KNN)[3]. K-Nearest Neighbor (KNN) is an algorithm used to classify data by utilizing training records from the nearest neighbors, where k is the number of closest neighbors used in the classification process [13]. KNN is a simple type of algorithm compared to other algorithms in machine learning. The working principle is to calculate the distance between the test data and the training data in the system, then look for the smallest or most similar value to classify the data. An object is classified with the majority class of the neighboring class, where the class that appears the most in a certain K constraint of its neighbor is taken, so that a new class is obtained according to the most common neighbor, if $K = 1$, then the object is assigned according to the class of its nearest neighbor. KNN is a supervised machine learning algorithm used for classification and regression, manipulating training data and classifying testing data based on distance metrics. [4] In the classification phase, K is a user-defined constant. The new class of a test data is classified by assigning the class that appears most frequently among the sample to the existing training data point. The K-Nearest Neighbor method in its working principle finds the closest distance between the data to be evaluated and the nearest neighbor K in the sample data. Euclidean distance is the distance between two points or coordinates calculated using the Pythagorean formula. It is the length of a straight line that connects two points in space, called point a and point b . This line, also known as a slash, runs between the x -axis and the y -axis with the coordinates given for point a and point b [7]. The following figure shows the calculation formula to find the distance between two points, namely the point on the training data (x) and the point on the testing data (y), so the Euclidean formula is used, as in the following equation.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Where D is the distance between the point on the training data x and the point of the testing data to be classified, where $x = x_1, x_2, \dots, x_i$ and $y = y_1, y_2, \dots, y_i$ and i present the value of the attribute and n is the dimension of the attribute.

Steps to calculate the KNN Algorithm method:

1. Specifies the parameter K (number of closest neighbors).
2. Calculate the square of the Euclid distance (query instance) of each object against the sample data that
3. Given
4. Then sort the objects into the group that has the smallest Euclidean distance.
5. Collecting category Y (classification of nearest neighbor)
6. By using the Nearest Neighbor category, which is the most majority, it is possible to predict the value of the query instance that has been calculated.

2.3. KIP Lecture

KIP Lecture is a program provided by the Indonesian government to support higher education for students who are financially disadvantaged. This program provides assistance with educational expenses, including tuition fees and living expenses, to students who meet certain requirements. KIP Lecture aims to ensure that every Indonesian citizen, regardless of economic background, has the same opportunity to continue their education at the university level. PIP is assistance in the form of cash, expansion of access, and learning opportunities from the government given to students and students from poor/vulnerable poor families to finance education. This is the basis for the government's commitment to placing access to higher education for all people as one of the development priorities.

PIP Higher Education for students is given in the form of the Indonesia Smart Lecture Card or KIP Lecture. PIP is intended for students who are accepted into higher education including people with disabilities with priority targets for high school KIP holders, students from poor/vulnerable poor families and/or with special considerations, affirmation students (Papua Region, 3T and TKI Children) as well as students affected by disasters, social conflicts or special conditions. As proof of the state's presence in helping people gain access to and guarantee higher education financing, since 2020 the government has provided KIP Lectures to more than 900,000 students who have been accepted into universities, including people with disabilities.

The main benefit of the 2024 Independent Lecture KIP is the guarantee of education fees paid directly to universities based on Study Program Accreditation (Prodi). All universities that receive KIP Lecture Merdeka students must also be officially accredited and recorded in the national university accreditation system. In addition, students who receive KIP Merdeka Lectures are also given living expense assistance. Monthly living expense assistance is given to students based on 5 regional clusters, namely IDR 800,000, IDR 950,000, IDR 1,100,000, IDR 1,250,000, and IDR 1,400,000 which refers to the results of a survey by the Central Statistics Agency. The living expense assistance is fully the student's right so that it is transferred directly to the recipient's student account. Students can take advantage of this assistance to meet various needs during college and should not be used by universities for any additional costs. Universities, LLDIKTI, or other parties may not utilize, use, or take the living expenses of all KIP Lecture recipients either through savings account books and/or ATMs, including not being allowed to keep savings account books and ATMs for the living expenses of KIP Lecture recipients[5].

2.4. Pemrograman Python

Python is an interpreter programming language and can be done with an object-oriented programming paradigm, functions, or in the usual way i.e. procedurally oriented programming, the application of python is used in making lightning analysis on transmission channels using an object-oriented programming paradigm[9],[11].

3. Results And Discucssion

The KNN algorithm must have data that is processed as a computational analysis to find prediction values where the data is in the form of test data and training data, the following is the analysis data that will be carried out in the search calculation process. application of the k-nearest neighbor (KNN) algorithm in machine learning to predict the selection of undergraduate study programs based on new KIP college students. The following data is as follows:

Table 1: Lecture KIP Assessment Dataset

No	Name	USM Values	Interview values	P3KE	Study Program Selection
1	Muhammad Rasya	580	70	5	2
2	Muhammad Alif Alfarino	550	80	0	2
3	M Yusri Hafizd	530	60	2	1
4	Hamdan Prayoga	500	93	5	1
5	Reynold Satria Mahendra	500	70	1	2
6	Putri Shyfa Khairani	500	90	0	1
7	Muhammad Zaki	500	90	0	2
8	Tiara Fransisca Br Sihole	470	75	0	2
9	Siti Fatimah	450	70	1	1
10	Mhd. Ahyar Faturrahim	450	91	2	1
11	Renita Br Tarigan	420	75	1	1
12	Doni Dwi Sutrisno	420	70	3	1
13	Novia Auliani	400	74	3	1
14	Ridho Hernanda	400	80	0	1
15	Dimas Syahputra	390	60	0	1
16	Putri Ramadhani	390	55	1	1
17	Sepri Herlinawati Hulu	390	90	1	2
18	Rohit Setiawan	380	60	1	1
19	Erly Putry Zefani	380	85	4	1
20	Rizky Pratama	370	65	0	2
21	Abdi Rahmadi Pohan	370	80	0	2
22	Piki Finelis Zebua	370	85	6	2
23	LIANA Sundari	340	92	0	2
24	Trivinni Mahulae	340	85	5	2
25	Dwi Ananda Sabrina	340	65	4	1
26	Muhammad Irham Lubis	320	70	4	1

27	Isna Zulaikha	320	80	3	1
28	Arief Alfahri	320	98	0	1
29	Christian Tambunan	310	90	1	2
30	Rahimah Faizah	310	85	4	1
31	Khairunnisa	310	75	0	1
32	Alputri Andinita	300	91	2	1
33	Jesika Zuina Br Lase	300	65	4	1
34	Joinadi Tama Buaton	300	75	4	1
35	Desti Imera Br Tarigan	290	87	0	1
36	Nita Rahayu	290	65	3	1
37	Suci Ayu Putri	290	96	3	1
38	Puza Rahayu	280	60	3	1
39	M. Ragib As Shiddiqi	280	94	4	1
40	Riko Jayadi	280	79	0	2
41	Annika Thalia Nonifaeri Lombu	280	88	3	2
42	Ayu Wulandari Br Ginting	280	88	4	1
43	Rasmawatita Br Sitepu	270	85	1	1
44	Dila Amanda	270	60	4	1
45	Aldo Arihdo Pradana	260	70	0	2
46	Grace Julietta Zendrato	260	70	2	1
47	Syarika Sahara Br Surbakti	250	75	7	1
48	Dwy Jayantika	250	65	0	1
49	Kasih Puri Rahayu	250	96	6	1
50	Della Aulia Prastia	250	95	5	1
51	Angel Agustina Siagian	230	60	0	1
52	Sandrina Elfira Damayanti	230	84	0	1
53	Novita	220	60	3	1
54	Ade Zulhiany Nasution	210	70	2	1
55	Windi Amelia	210	70	3	1
56	Arfiva Syahbina Ashadi	200	60	0	2
57	Patia Azzahra	190	75	4	1
58	Adisti Pehulisa Br Sembiring	170	60	0	2
59	Lamtiur Silaban	170	85	0	1

In the explanation of the table above, there are the names of the students and the assessment carried out by the campus in the selection of the KIP Lecture, the results of the study program taken by students with the number 1 is the information systems study program and the number 2 is the informatics engineering study program and P3KA is the poverty level from decile 1 to decile 7 which is data by the social service.

4. Testing

The results of the program display analysis process predicting the application of the k-nearest neighbor (KNN) algorithm in machine learning to predict the selection of undergraduate study programs based on new students can be seen in the figure below:

```
In [15]: # Data baru untuk diprediksi
new_data_csv = [[400, 90, 1]]

In [16]: # Membuat model KNN
knn_csv = KNeighborsClassifier(n_neighbors=3) # k=3
knn_csv.fit(X, y) # Melatih model

Out[16]:
KNeighborsClassifier
KNeighborsClassifier(n_neighbors=3)

In [24]: # Prediksi data baru
prediction_csv = knn_csv.predict(new_data_csv)
class_label_csv = "Sistem Informasi" if prediction_csv[0] == 0 else "Teknik Info
```

Fig 1: KNN Program Test Results

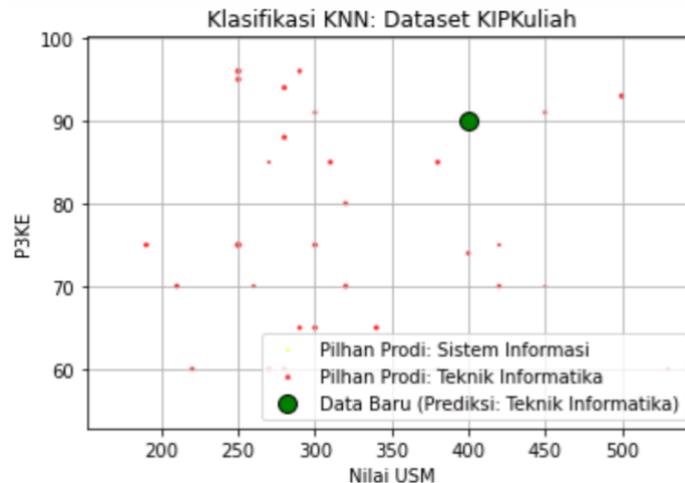


Fig. 2: KNN Prediction Graph Results

It can be concluded that the results of the new data prediction with USM Score 400, Interview Score 40 and P3K3 Poverty Level Decile 1 new KIP Lecture students who will be included in STMIK Kaputama in the selection of the chosen study program is Informatics Engineering.

5. Conclusion

The KNN algorithm can solve the problem of predicting the application of the k-nearest neighbor (KNN) algorithm in machine learning to predict the selection of undergraduate study programs based on new students. Based on the test, 59 existing datasets were used with a prediction value of $K = 3$. The level of new KIP Lecture students who will enter STMIK Kaputama in the selection of the chosen study program is Informatics Engineering, as evidenced by testing the Python programming language system.

References

- [1] Y. Saputra, D. Jaelani, and E. S. Nurpajriah, "Implementasi Algoritma Smart Untuk Beasiswa Kip-K Di Perguruan Tinggi (Studi Kasus: Uin Sunan Gunung Djati Bandung) Implementation of the Smart Algorithm for Kip-K Scholarships in Higher Education (Case Study: State Islamic University Sunan Gunung Djati Bandung)," *J. Sist. Inf. Dan Bisnis Cerdas*, vol. 17, no. 1, p. 59, 2024.
- [2] S. Jesus, P. Saleiro, B. M. Jorge, R. P. Ribeiro, and R. Ghani, "Aequitas Flow : Streamlining Fair ML Experimentation," vol. 25, pp. 1–8, 2024.
- [3] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.
- [4] F. Malik Namus Akbar, "Metode KNN (K-Nearest Neighbor) untuk Menentukan Kualitas Air," *J. Tekno Kompak*, vol. 18, no. 1, pp. 28–40, 2024.
- [5] Kemdikbud, "Pedoman Pendaftaran KIP Kuliah Merdeka 2024," *Kementerian Pendidik. dan Kebud. Republik Indones.*, p. 23, 2024, [Online]. Available: <https://lldikti6.kemdikbud.go.id/wp-content/uploads/2022/08/PUSLAPDIK-20220725-Bahan-Pendampingan-KIPK-LLDIKTI-PTS.pdf>
- [6] Wijoyo A, Saputra A, Ristanti S, Sya'ban S, Amalia M, and Febriansyah R, "Pembelajaran Machine Learning," *OKTAL (Jurnal Ilmu Komput. dan Sci.)*, vol. 3, no. 2, pp. 375–380, 2024, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal/article/view/2305>
- [7] M. Rahmadiyah and P. Suparman, "Penerapan Metode K-Nearest Neighbour Untuk Sistem Penentuan Peminjaman Modal Nasabah Bank Syariah Indonesia Cabang Cikarang Berbasis Website," *J. Inf. dan Komput.*, vol. 10, no. 2, pp. 189–197, 2022.
- [8] G. Saputro, K. Aurora, T. Winarko, A. Shayla, and A. Saifudin, "Penggunaan Machine Learning untuk Memprediksi Defect pada Pengembangan Perangkat Lunak," vol. 2, no. 2, pp. 300–305, 2024, [Online]. Available: <https://jurnalmahasiswa.com/index.php/biikma>
- [9] W. F. Mustafa, S. Hidayat, and D. H. Fudholi, "Prediksi Retensi Pengguna Baru Shopee Menggunakan Machine Learning," *J. Media Inform. Budidarma*, vol. 8, no. 1, p. 612, 2024, doi: 10.30865/mib.v8i1.7074.
- [10] Z. Khairina, M. Simanjuntak, and J. N. Sitompul, "Sistem Pendukung Keputusan Penerimaan Kartu Indonesia Pintar (KIP) Pada Siswa Menggunakan Metode Moora," *Semin. Nas. Inform.*, pp. 12–20, 2021.
- [11] M. Reza et al., "Artificial Intelligence : Image Processing & Application with Python," *Semin. Nas. Pengabd. Masy. LPPM UMJ*, vol. 1, no. 1, pp. 1–8, 2022, [Online]. Available: <http://jurnal.umj.ac.id/index.php/semnaskat>.