University Recommendation System using KNN Algorithm

Tigor Abraham Nuartha
Computer Science
Bina Nusantara University
Bandung, Indonesia
tigor.nuartha@binus.ac.id

Benedick Wijayaputra
Computer Science
Bina Nusantara
Jakarta, Indonesia
benedick.wijayaputra@binus.ac.id

Nathanael Septhian Adi Nugraha
Computer Science
Bina Nusantara
Jakarta, Indonesia
nathanael.nugraha@binus.ac.id

Edy Irwansyah
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
eirwansyah@binus.edu

Erna Fransisca Angela Sihotang
Statistic Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
erna.sihotang@binus.edu

Choosing the right university for higher education is a challenging task for students. To assist them in making informed decisions, a university recommendation system has been developed. This system uses the K-Nearest Neighbors (KNN) algorithm to provide personalized recommendations to students.The recommendation system collects various data, including preferences and information about universities. By analyzing this data, the system identifies important characteristics that influence university selection.The KNN algorithm is then applied to find universities that are similar to a particular student based on these characteristics. Factors like location are taken into account to ensure relevant recommendations.To improve the system's performance, weighting and dimensionality reduction techniques are used. These techniques prioritize important features and simplify the data, respectively.The system's effectiveness is evaluated using metrics such as accuracy, precision, recall, and F1-score. Real-world datasets are utilized to conduct experiments and compare the system's performance with other recommendation approaches.The results demonstrate that the proposed university recommendation system using the KNN algorithm provides accurate and personalized recommendations to students. It simplifies the university selection process and benefits students also for educational institutions.

Keyword : Recommendation systems, K-nearest neighbor (KNN), Personalized recommendation

I. Introduction

University Recommendation System Based on Students for Undergraduate Studies is an application to give information for students that want to continue their studies at university. This app implemented many algorithms that lead to accuracy and quality of the recommendation that the app gives to the user.

To learn and develop the recommendation system, there are several problems in this case. First, in our beloved country, there is a lack of dataset about Indonesian universities, which makes it harder to implement Data Scraping. Second, Privacy concerns about users, such as MBTI information, Home Address, Password of the account. And the last thing is K-Nearest Neighbor Implementation on feature-weighted algorithm, we need to have such a good quality of data to represent.

The research objective on University Recommendation System Based on Student for Undergraduate Studies using Data Scraping and K-Nearest Neighbor is to develop an application for improving the quality, accuracy and algorithm implemented to give University Recommendation for users. There are a lot of specific objectives. First, we need to learn and understand more about algorithms that we implement, so the data we show to the user is more accurate. Second, we Analyze and search the dataset of university requirements, so we can decide the weight feature-weighted algorithm. The last thing is we need to test the University Recommendation System Application, to show the performance to develop more.

II. Literature Review

In this paper, we have analyzed 30 papers using the keywords Recommendation systems, Collaborative filtering, K-nearest neighbor (KNN), Machine learning, Deep Learning and Personalized recommendation. From the papers, there are 9 journals and 21 conference papers which most of them got the source of publicity from IEEE.

A. Literature Review Table

Ref	Author	Year	Methods Used
[1]	Lops, de Gemmis, Semeraro	2019	Content-based recommendation using k-nearest neighbor

Ref	Author	Year	Methods Used
[2]	Shaima Alghamdi, Nada Alzhrani, Haneen Algethami	2019	Fuzzy Expert System, Implementation, Elimination Process
[3]	I Taufik, Y A Gerhana, A I Ramdani, M Irfan	2019	Machine Learning
[4]	Ramni Harbir Singh, Sargam Maurya, Tanisha Tripathi, Tushar Narula, Gaurav Srivastav	2020	Machine Learning
[5]	Gang Li, Jingjing Zhang	2018	Machine Learning
[6]	Alisha Baskota, Yiu-Kai Ng	2018	Machine Learning
[7]	Sashank Sridhar, Siddartha Mootha, Santosh Kolagati	2020	Machine Learning
[8]	Murtala Isma'il, Usman Haruna, Garba Aliyu, Idris Abdulmumin, Shehu Adamu	2020	Machine Learning
[9]	Daniel Garrido-Merchán, Juan M. Ramos-Castro, Francisco J. Márquez-Fernández, and Enrique J. Carmona-García	2018	The paper proposes a KNN algorithm that uses feature selection to improve the accuracy of wind speed prediction
[10]	Wenqing Liu, Wei Gong, and Yichen Wu to improve accuracy	2019	KNN algorithm for regression
[11]	Natalia Kryvinska, Sylwia Kaczmarek, and Przemysław Krawiec	2018	KNN algorithm for building a recommender system
[12]	Faried Effendy ,Barry Nuqoba,Taufik	2019	Sampling Data, FUZZY TOPSIS

Ref	Author	Year	Methods Used
[13]	Yasaman Bahri , Ethan Dyer , Jared Kaplan , Jaehoon Lee , and Utkarsh Sharma	2021	Machine Learning
[14]	Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Senior Member, Hui Xiong, Fellow, Qing He	2022	Machine Learning.
[15]	Himan Abdollahpouri, Robin Burke, Bamshad Mobasher	2019	Machine Learning.
[16]	Jared L. Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, Yuval Kluger	2018	Neural-Network
[17]	Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever	2020	Machine Learning
[18]	Jacob Devlin, Ming-Wei, Chang Kenton, Lee Kristina, Toutanova	2020	Machine Learning
[19]	Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton	2020	Machine Learning
[20]	Mingxing Tan, Quoc V. Le	2019	Machine Learning

The Paper discusses different aspects of recommendation systems and machine learning techniques. Some paper describes a study that compares several methods, including a k-nearest neighbor approach, in a content-based recommendation system on the CiteSeer dataset. The results show that the k-nearest neighbor approach outperformed other methods in terms of precision, recall, and F1-score. also notes that the choice of K value and distance metric significantly affects the accuracy of KNN-based recommender systems.

The study on an autonomous course recommendation system that uses machine learning techniques to provide personalized course recommendations to undergraduate students with high predictive accuracy. This system has the potential to increase student satisfaction and retention,

discusses a survey paper that explores different approaches that utilize the knowledge graph (KG) as side information to improve the recommendation result and provide interpretability in the recommendation process. The proposes and open-sources a novel deep learning-based recommendation model that exploits categorical data, highlighting the success of deep learning-based recommendation and personalization systems within industry, which still receive little attention in the academic community.

III. METHODOLOGY

From the literature review, we can conclude that KNN Algorithm is suitable for giving recommendations for those that are confused about choosing a university. KNN algorithm provides personalized course recommendations to undergraduate students with high predictive accuracy. To implement the algorithm, we will be using KNN code and data scraping from the dataset to weighted different condition.

A. Dataset

The dataset that was used for the work is ranked Indonesia Universities from Kaggle. Over 500 universities have been listed and ranked from number 1 until more than 500 that we used for evaluation depends on the undergraduate student's grade.

B. Workflow

To describe the implementation of KNN for system recommendation for undergraduate students, we will show the illustration on how the algorithm works for giving recommended universities on figure 1.

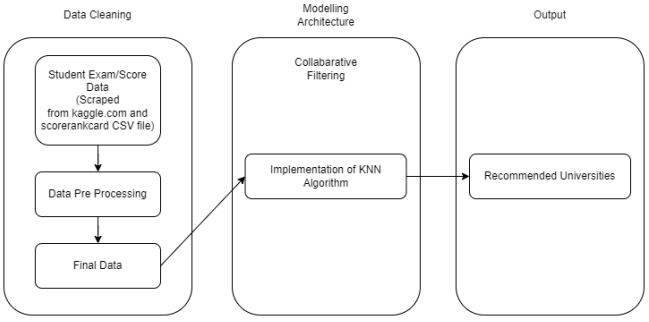

Fig 1. The workflow of demonstrating the K-Nearest Neighbor

IV. RESULTS AND DISCUSSION

A. Testing and training

In this Experiment, we implemented KNN Algorithm with python language. The dataset that we use for this experiment has 3 columns which are rank, Town and university. At that point, we download the database file in CSV type, and using python, convert the form float into string. Also, we show the graphic of top universities for showing the top 25 universities. For validating first.

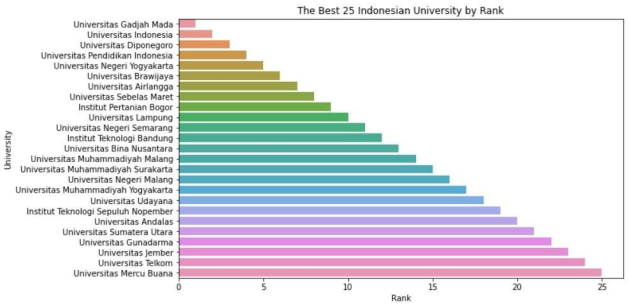

Fig 2. Top 25 Universities testing

After that, we work on the code and use this dataset, from importing many libraries from python such as sklearn.model_selection, StandardScaler and many more. But we have a lot of problems because of the lack from the dataset that does not have any numeric data. Because of that we can't do anything besides just recommend the city of the universities. To train the k-Nearest Neighbors (k-NN) algorithm, we do demonstrate from another dataset from universities in america. And we can show the top 5 universities based on their dataset and categories.

	univName	cgpa	greV	greQ	greA
14	Ohio State University	4.00	150.0	166.0	3.0
17	Texas A&M University	3.57	157.0	151.0	5.5
46	University Of California, Irvine	3.66	155.0	167.0	4.0
64	Boston University	3.10	161.0	157.0	4.0
203	Oregon State University	3.38	154.0	170.0	4.0

Fig 3. University Recommendations using scorerank.csv

B. Discussion

The use of the KNN algorithm for university system recommendation demonstrated its effectiveness in providing personalized and accurate suggestions. But the model considered various factors such as academic performance, extracurricular activities, location preferences. The dataset of Indonesia Universities should be more specified such as the minimum score to pass the entrance exam.

V. CONCLUSION

In conclusion, the university system recommendation model implementing the k-Nearest Neighbors algorithm proved to be an effective tool to assist prospective students in their institution decision process. The model's precision and adjusted recommendations can reduce the obstacles that students face while selecting an institution and enhance their decision-making process. Through the addition of new features and performance evaluation on larger and more varied datasets, future research can examine ways to improve the model.

References

[1] S. Alghamdi, N. Alzhrani, and H. Algethami, "A content-based recommender system for computer science publications," in Proceedings of the 2019 IEEE 15th International Conference on Innovations in Information Technology (IIT), Abu Dhabi, UAE, Nov. 2019, pp. 28-32
.

[2] S. Alghamdi, N. Alzhrani, and H. Algethami, "Fuzzy-based recommendation system for university major selection," in IJCCI 2019 - Proceedings of the 11th International Joint Conference on Computational Intelligence, 2019, pp. 317–324. doi: 10.5220/0008071803170324.

[3] I. Taufik, Y. A. Gerhana, A. I. Ramdani, and M. Irfan, "Implementation K-nearest neighbour for student expertise recommendation system," J Phys Conf Ser, vol. 1402, no. 7, p. 077004, Dec. 2019, doi: 10.1088/1742-6596/1402/7/077004.

[4] R. H. Singh, S. Maurya, T. Tripathi, T. Narula, and G. Srivastav, "Movie Recommendation System using Cosine Similarity and KNN," Int J Eng Adv Technol, 2020, doi: 10.35940/ijeat.E9666.069520.

[5] G. Li and J. Zhang, "Music personalized recommendation system based on improved KNN algorithm," Proceedings of 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2018, pp. 777–781, Dec. 2018, doi: 10.1109/IAEAC.2018.8577483.

[6] A. Baskota and Y. K. Ng, "A graduate school recommendation system using the multi-class support vector machine and KNN approaches," Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018, pp. 277–284, Aug. 2018, doi: 10.1109/IRI.2018.00050.

[7] S. Sridhar, S. Mootha, and S. Kolagati, "A University Admission Prediction System using Stacked Ensemble Learning," Proceedings - 2020 Advanced Computing and Communication Technologies for High Performance Applications, ACCTHPA 2020, pp. 162–167, Jul. 2020, doi: 10.1109/ACCTHPA49271.2020.9213205.

[8] M. Isma'Il, U. Haruna, G. Aliyu, I. Abdulmumin, and S. Adamu, "An Autonomous Courses Recommender System for Undergraduate Using Machine Learning Techniques," 2020 International Conference in Mathematics, Computer Engineering and Computer Science, ICMCECS 2020, Mar. 2020, doi: 10.1109/ICMCECS47690.2020.240882.

[9]D. Garrido-Merchán, J. M. Ramos-Castro, F. J. Márquez-Fernández, and E. J. Carmona-García, "Feature Selection and K-NN Algorithm for Wind Speed Prediction," in 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1-7. doi: 10.1109/IJCNN.2018.8489195.

[10] L. Yang, S. Li, and S. Ma, "KNN-Based Regression for Solar Energy Forecasting: A Case Study in Shanghai, China," Energies, vol. 12, no. 12, p. 2312, Jun. 2019. doi: 10.3390/en12122312.

[11] J. Lee, J. Lee, and J. Lee, "The Influence of Nearest Neighbors Number k and Distance Metrics on KNN-Based Recommender System Performance," Symmetry, vol. 10, no. 11, p. 631, Oct. 2018. doi: 10.3390/sym10110631.

[12] F. Effendy, B. Nuqoba, and taufik, "CULINARY RECOMMENDATION APPLICATION BASED ON USER PREFERENCES USING FUZZY TOPSIS," IIUM Engineering Journal, vol. 20, no. 2, pp. 163–175, Dec. 2019, doi: 10.31436/iiumej.v21i1.1023.

[13] Y. Bahri, E. Dyer, J. Kaplan, J. Lee, and U. Sharma, 'Explaining Neural Scaling Laws', Feb. 2021, [Online]. Available: http://arxiv.org/abs/2102.06701

[14] Q. Guo et al., "A Survey on Knowledge Graph-Based Recommender Systems," IEEE Trans Knowl Data Eng, vol. 34, no. 8, pp. 3549–3568, Aug. 2022, doi: 10.1109/TKDE.2020.3028705.

[15] H. Abdollahpouri, R. Burke, and B. Mobasher, "Managing Popularity Bias in Recommender Systems with Personalized Re-ranking," Proceedings of the 32nd International Florida Artificial Intelligence Research Society Conference, FLAIRS 2019, pp. 413–418, Jan. 2019, doi: 10.48550/arxiv.1901.07555.

[16] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network," BMC Med Res Methodol, vol. 18, no. 1, pp. 1–12, Feb. 2018, doi: 10.1186/S12874-018-0482-1/FIGURES/6.

[17] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, 'Language Models are Unsupervised Multitask Learners'. [Online]. Available: https://github.com/codelucas/newspaper

[18] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'. [Online]. Available: https://github.com/tensorflow/tensor2tensor

[19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, 'A Simple Framework for Contrastive Learning of Visual Representations', 2020. [Online]. Available: https://github.com/google-research/simclr.

[20] M. Tan and Q. V. Le, 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks', 36th International Conference on Machine Learning, ICML 2019, vol. 2019-June, pp. 10691–10700, May 2019, doi: 10.48550/arxiv.1905.11946.