

Machine learning based Expert System for Online Admissions

Disha Wankhede¹, Sanika Sasane², Aditi Igade³, Anish Shet⁴ and Vighnesh Kandalgaonkar⁵

¹⁻⁵Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune

Email: disha.wankhede.@viit.ac.in, {sanikasasane2003, aditiigade20, ,anishshet5,vkandalgavkar850 }@gmail.com

Abstract— The web application permits on-line admissions saving the time of geographically scattered college students .It permits lowering time in activities, centralized information dealing with and paperless admission with decreased manpower .It improves operational performance and decreases the fee. The college admission device has been advanced so as to automate admission device beginning from the notification to admission procedure .The device used a selection tree primarily based totally category method. Prior to device improvement, six fashions have been created and examined to discover the maximum green version which might later be carried out for improvement of internet-primarily based totally pupil admission screening device. Online Admission screening device is an internet device that may be accessed at some stage in the company and out of doors as nicely with right login furnished. The purpose of this studies paintings is to significantly verify the gathering and processing of pupil admission shape procedure so as to become aware of troubles encountered with inside the device, a good way to layout a brand new device a good way to assist lessen or absolutely remove the diagnosed issues internet. The automatic admission processing device will assist lessen a number of the onerous paintings contain in pupil admission processing, maximized returns for the group offer a pleasant consumer interface users.

Index Terms— Admission Process, Machine Learning.

I. INTRODUCTION

Admission of undergraduate students to Thai educational institutions is very important because it directly affects education management, budget planning for administration and instruction, and, lastly, the standard indicator of each university that is largely focused on students and the quality of education. Because of this, the educational institutions put a lot of effort into helping students enroll in programs and stay enrolled until they complete their studies in a high-quality manner within the allotted time [1]. The Department of Science of Suratthani Rajabhat University and Technology consistently receives a large number of applications, and there are several options for new students to apply. However, a number of staff members are needed for each student admission in order to assess the applicant profile and select the most qualified candidates based on departmental requirements[2]. This research is all about helping students become better people through the Internet. It's using data mining to create an online entrance exam system that helps reduce evaluation time and staff [3]. It also helps applicants pick the right career based on their skills and achievements. The system was created by analyzing students' profiles to come up with 6 decisions. It's an efficient and popular way of using data mining. The six models are made up of ambiguous

techniques used for modeling. One of the trio of models uses a K-fold cross validation technique, while the other three use percentages. We then compared the models to see which one is the most effective for creating an online student entrance exam system. Not only does it save time, but it also helps students choose the right field of study, which is best suited to their characteristics and the university's goals.

An information device to as it should be forecast pupil admission through the internet. The system makes use of seven forecast fashions, along with k-fold cross-validation, percent split, and statistics separation techniques[4]. The separation records approach outperforms all others, with an accuracy of 94%, precision of 94.30% and F-measure of 93.70%. The gadget was evaluated by means of experts and personnel, with a mean of 4.17 and 4.34 respectively, indicating fine performance[5]. Data mining strategies to expand a version for deciding on high college programs, specializing in factors influencing program choice and college students' ability to complete the chosen software. The look at used information from 850 Satri Si Suriyothai college students in 2012, revealing that basic subjects rankings, GPA, and grade point average directly have an effect on software choice and success[6]. Data mining strategies to predict schooling achievement of Ratchaphruek college students. The version makes use of non-public information and registration statistics from four specializations: advertising, enterprise computer, management, and lodge and Tourism control. The outcomes display that core finance problem organization impacts business computer and inn and Tourism management greater, whilst core commercial enterprise concern institution influences marketing and control extra. The prediction models have excessive accuracy prices.[7] Decision tree algorithms to provide college students with pointers for specialization choice, addressing the lack of information and experience in choosing a specialization. The examine found that these algorithms are highly green, with accuracy charges exceeding 80%.[8] The use of statistics mining to research candidate profiles and shop consequences in a database for future student admissions[17]. This research concludes two modules: studying specialization choice behavior using ok-manner clustering, attempting to find affiliation policies the usage of an Apriori algorithm, and evaluating two fashions for forecasting new college students the usage of choice tree and multilayer notion-based totally synthetic neural community models.[9][10] Different machine learning and deep learning models applied on healthcare sector for experiment purpose and got good accuracy using model.[11][12][16] Framework required cloud space and secure communication as mentioned by author in there paper .[13][14][15]

II. PROPOSED METHODOLOGY

a) Need of Proposed System

There are various shortcomings in the existing system that should be over come such as lot of manual work for the administrative staff, huge paperwork in filling out the forms and application, waste of time standing in queue for hours, the need to physically visit banks to pay the fees and then resubmit the receipts to the institution for confirmation, etc. All these problems can be solved if the system gets automated with automation system can achieve:-

- 1) All the paperwork will be in digital format.
- 2) No need to visit the institutions as the process will be automatized.
- 3) Faster admission Process.
- 4) Online Document Verification.
- 5) Automatic allocating Unique ID to each student.
- 6) Parents and Students both will be updated timely with the latest notices.

As shown in following Figure 1, 2 and 3.

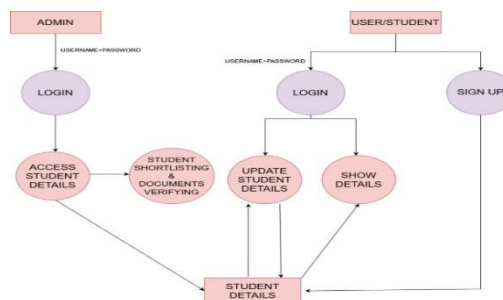


Fig1.Proposed system

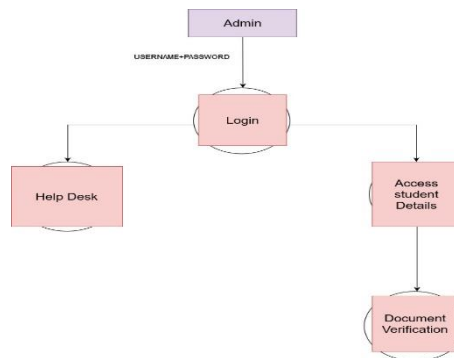


Fig 2. Framework Admin Side

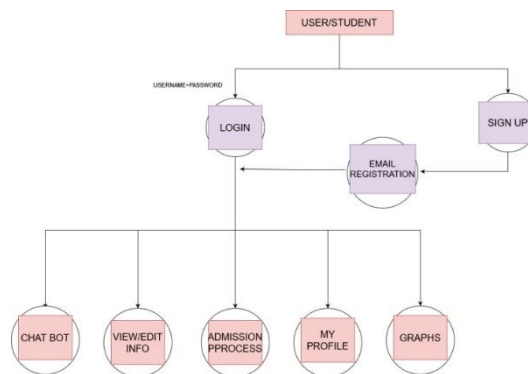


Fig 3-Framework User Side

III. METHODS OF DATA MINING ALGORITHMS

i) *Decision Tree-Based Algorithm*

The Decision Tree algorithm, a prominent method in the field of machine learning, has received significant attention due to its versatility and ability to provide clear interpretations in both classification and regression tasks. By utilizing a model that resembles a tree, the algorithm systematically divides data based on the most distinguishing features, creating a well-organized and easy-to-understand framework for decision-making. The algorithm's capacity to handle both numerical and categorical data is one of its features, which makes it appropriate for a variety of real-world scenarios. Practitioners should be cautious when using overfitting models, even if they work well on training data. Overfitting is observed when a model excels in accurately predicting the training data, but faces challenges in making accurate predictions for new cases. Techniques such as pruning and controlling the maximum depth of the tree are commonly employed to address this challenge. As decision trees continue to hold a crucial position in the field of machine learning, ongoing research endeavors to enhance their resilience, scalability, and adaptability in various domains.

ii) *K-fold Cross Validation Technique*

K-fold Cross Validation (CV) is a crucial technique in the field of machine learning model evaluation. It provides a comprehensive approach to assess the performance and generalizability of a model. This method involves dividing the dataset into k folds of equal size. During each iteration, k-1 folds are utilized for training the model, while the remaining fold is employed for validation purposes. To make sure that every data point is used as a training and validation instance, this process is repeated k times. By averaging the performance metrics across all folds, K-fold CV offers a more robust and reliable estimation of a model's effectiveness, reducing the impact of dataset variability. This technique is particularly valuable when working with limited data, as it maximizes the utilization of available information and aids in identifying potential issues such as overfitting. As machine learning models become more complex and datasets become more diverse, the application and enhancement of K-fold CV

continue to be a focal point of research, making significant contributions to the advancement of model evaluation methodologies.

iii) Percentage Split

The Percentage Split Technique, which is a fundamental methodology in the evaluation of machine learning models, entails the division of a dataset into training and testing sets based on a specified percentage split. This technique simplifies the process by allocating a certain percentage, typically ranging from 70% to 80%, of the data for training the model, while reserving the remaining portion for evaluating its performance. Although this approach is straightforward, it does possess certain limitations, particularly in cases where datasets are restricted or imbalanced. The selection of the percentage split is of utmost importance, as an inadequate allocation may result in overfitting or underfitting complications. Despite its simplicity, the Percentage Split Technique serves as a foundational basis for model evaluation, especially in situations where computational resources are constrained or during the initial stages of model development. Ongoing research endeavors are exploring variations and optimizations of this technique to enhance its applicability across diverse datasets and ensure its relevance in the ever-evolving landscape of machine learning methodology.

IV. RESULTS AND DISCUSSION

The WEKA software was employed to fabricate models. Subsequently, these models were juxtaposed in terms of accuracy, precision, recall, and F-measure to discern the most efficacious one for the construction of an internet-based student admission screening system. The performance of each model is shown in the table below.

TABLE I: PERFORMANCE OF EACH MODEL

Modeling	Times (Seconds)	Precision (%)	Recall (%)	F-Measure (%)	Accuracy (%)
Cross Validation (5 folds)	0.11	88.00	88.50	87.40	87.50
Cross Validation (10 folds)	0.02	87.90	87.40	87.30	87.40
Cross Validation (100 folds)	0.02	88.10	87.60	87.50	87.60
Cross Validation (90 folds)	0.00	88.50	87.80	87.50	87.76
Cross Validation (80 folds)	0.10	87.90	87.80	87.60	87.82
Cross Validation (70 folds)	0.00	87.40	87.10	86.90	87.12

To gauge the accuracy of a specific model, it is necessary to evaluate each class. **Precision** can be calculated by dividing true positives (TP) by the sum of true positives and false positives (FP).

$$Precision = TP / (TP + FP)$$

Similarly, Recall can be determined by dividing TP by the sum of TP and false negatives (FN).

$$Recall = TP / (TP + FN)$$

F-measure, which takes into account both precision and recall, can be calculated by using the formula $(2 \times Precision \times Recall) / (Precision + Recall)$.

$$F\text{-measure} = (2 \times Precision \times Recall) / (Precision + Recall)$$

Accuracy, on the other hand, measures the overall integrity of a model by considering all classes.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

Modeling was conducted utilizing methods based on decision trees. The six models were assessed in relation to accuracy, precision, recall, and F-measure. A visual representation of the performance comparison was presented in Figure 4.

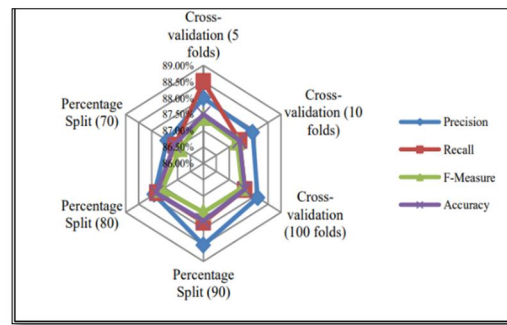


Fig. 4 Various split representation using graphs

Six models with varying techniques were used in a classification experiment. The precision, recall, F-measure, and accuracy of each model were compared, and the results showed that all six models produced results similar to Table 1. Experiment results showed that, when all factors were taken into account, the model with the highest precision (as expressed in percentage) was 90 Percentage Split (88.50), which was followed by Cross validation (100 folds) (88.10), Cross validation (5 folds) (88.00), Cross validation (10 folds) and Percentage Split (80) (both at 87.90), and Percentage Split (70) (27.30%). According to the experiment, the model with the highest percentage recall, or cross validation (5 folds), was percentage split (90), percentage split (80), and so on. (both at 87.80), Percentage Split (70) (87.10), Cross validation (10 folds) (87.40), and Cross validation (100 folds) (87.60), in that order. In terms of F-measure, which is expressed as a percentage, Percentage Split (80) (87.60) was the model with the highest F-measure. This was followed by Percentage Split (70) (86.90), Cross validation (5 folds) (87.40), Cross validation (10 folds) (87.30), and finally, Cross validation (100 folds), which yields the same F-measure as Percentage Split (90) (87.50). In summary, in terms of accuracy expressed as a percentage, Percentage Split (80) (87.82) was the model with the highest accuracy, ahead of Percentage Split (90) (87.76), Cross validation (100 folds) (87.60), Cross validation (5 folds) (87.50), Cross validation (10 folds) 87.40, and Percentage Split (70) (87.12).

V. CONCLUSION

In conclusion, the development and implementation of a Student Admission Screening System embody a significant advancement towards augmenting efficiency and transparency in the admission process. The assimilation of cutting-edge technologies and data-driven decision-making within the system not only simplifies the workflow for administrators but also furnishes applicants with a user-friendly interface. Through the automation of the screening process, the system diminishes the probability of errors, ensures uniformity, and expedites the evaluation of candidate qualifications. The instantaneous accessibility of information promotes prompt communication with applicants, thereby contributing to a more seamless and responsive admission experience. Furthermore, the system's analytical capabilities empower institutions to identify patterns, evaluate the efficacy of admission criteria, and continuously refine the screening process. As educational institutions navigate the dynamic landscape of admissions, the Student Admission Screening System stands as a valuable instrument, aligning technological innovation with the pursuit of academic excellence.

REFERENCES

- [1] OECD , "Making the education system more responsive to labour market needs" , OECD Skills Strategy Poland , 11 Dec 2019, doi : <https://doi.org/10.1787/b377fbcc-en>
- [2] Jorge Amar, Sohrab Rahimi, Nicolai von Bismarck and Akshar Wunnava , "Smart scheduling: How to solve workforce-planning challenges with AI" , November 1, 2022
- [3] Liming Huang , Tingting Zheng , Qiaomin Huang , "Online Labor Education Optimization Method Based on Computer Intelligent Algorithm", 2023 Aug 23 ,doi : 10.1155/2022/8740978 ,
- [4] Hongtao Tang, Bo Fang, Rong Liu , "A hybrid teaching and learning-based optimization algorithm for distributed sand casting job-shop scheduling problem " , 25 February 2019,doi : 10.1016/j.asoc.2022.108694
- [5] Shome, S., & Gupta, Y. (2020). A reflection of Indian higher education system: An academic labour market perspective. *Research in Education*, 106(1), March-21, doi : <https://doi.org/10.1177/0034523718793162>

- [6] G. Hu, "Exploration and Analysis of the Path Optimization of Labour Recommendation Model Based on Big Data and Auto-Kg Algorithm," 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 2022, pp. 783-786, doi : 10.1109/ICPECA53709.2022.9719269.
- [7] Liming Huang, Tingting Zheng, Qiaomin Huang " Online Labor Education Optimization Method Based on Computer Intelligent Algorithm " , Computational Intelligence and Neuroscience ,2022 Aug 17, doi :10.1155/2023/9853670
- [8] Chia-Huei ,Chen Wei ,Ren Daowen ,Pang Meiyan " Design of Online Education Information Management System Based on Data Mining Algorithm " , Wireless Communications and Mobile Computing , 2022/08/04, doi:10.1155/2022/6570922
- [9] Lauder Hugh ,Mayhew Ken , " Higher education and the labour market: an introduction " , Oxford Review of Education , 2020 January 2, doi : 10.1080/03054985.2019.1699714
- [10] Hongtao Tang, Bo Fang, Rong Liu , "A hybrid teaching and learning-based optimization algorithm for distributed sand casting job-shop scheduling problem" , Applied Soft Computing , Volume 120, May2022, doi:https://doi.org/10.1016/j.asoc.2022.108694
- [11] Mrs. Disha Sushant Wankhede, Dr. Selvarani Rangasamy, "REVIEW ON DEEP LEARNING APPROACH FOR BRAIN TUMOR GLIOMA ANALYSIS" Journal of Information Technology in Industry, VOL. 9 NO. 1 (2021) pp. 395 - 408 , DOI: https://doi.org/10.17762/itii.v9i1.144
- [12] Disha Sushant Wankhede, R. Selvarani, Dynamic architecture based deep learning approach for glioblastoma brain tumor survival prediction, Neuroscience Informatics, Volume 2, Issue 4, 2022, 100062, ISSN 2772-5286, https://doi.org/10.1016/j.neuri.2022.100062. (https://www.sciencedirect.com/science/article/pii/S2772528622000243)
- [13] G. S. Prasad and V. S. Gaikwad, "A survey on user awareness of cloud security", Int. J. of Engineering and Technology, vol. 7, no. 2.32, pp. 131-135, May 2018
- [14] Gudapati S.P., Gaikwad V. (2021) Light-Weight Key Establishment Mechanism for Secure Communication Between IoT Devices and Cloud. In: Satapathy S., Bhateja V., Janakiramaiah B., Chen YW. (eds) Intelligent System Design. Advances in Intelligent Systems and Computing, vol 1171. Springer, Singapore. https://doi.org/10.1007/978-981-15-5400-1_55
- [15] Shrimant, Gaikwad Vidya, Ravindranath, K. & Prasad, Gudapati Syam(2023) A mathematical model for secure Cloud-IoT communication: Introducing the revolutionary lightweight key mechanism, Journal of Discrete Mathematical Sciences and Cryptography, 26:5, 1341–1354, DOI: 10.47974/JDMSC-1750
- [16] Wankhede, D.S., Shelke, C.J. (2023). An Investigative Approach on the Prediction of Isocitrate Dehydrogenase (IDH1) Mutations and Co-deletion of 1p19q in Glioma Brain Tumors. In: Abraham, A., Pillana, S., Casalino, G., Ma, K., Bajaj, A. (eds) Intelligent Systems Design and Applications. ISDA 2022. Lecture Notes in Networks and Systems, vol 715. Springer, Cham. https://doi.org/10.1007/978-3-031-35507-3_19
- [17] Sutar, S. ., Jose, K. ., Gaikwad, V. ., Mishra, V. ., Wankhede, D. ., & Karnik, M. . (2023). Enhancing Data Management: An Integrated Solution for Database Backup, Recovery, Conversion, and Encryption Capabilities. International Journal of Intelligent Systems and Applications in Engineering, 12(6s), 720–734. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4011