CAREEX – AN AI ASSISTED CAREER GUIDANCE AND ELIGIBILITY PREDICTION SYSTEM

Dr Soma Prathibha
Department of Information Technology
Sri Sai Ram Engineering College
Chennai, 600044, India
prathibha.it@sairam.edu.in

G. Shamini
Department of Information Technology
Sri Sai Ram Engineering College
Chennai, 600044, India
shamini2522@gmail.com

Dharshini S
Department of Information Technology
Sri Sai Ram Engineering College
Chennai, 600044, India
sec20it060@sairamtap.edu.in

U. Rusitha
Department of Information Technology
Sri Sai Ram Engineering College
Chennai, 600044, India
rusithaudhaya@gmail.com

Abstract—The education sector will be seeing a vast change in the way it functions in this digital world. As we shut our doors to the outside, many options and other alternatives have opened right inside our house. There will be two types of people choosing their career. 1. Individuals are stuck as 'What next'. Since the lack of guidance and not wanting to study what your parents or relatives want us to, has sucked us all many times. 2. There are some individuals who want to enter a particular university, but don't know how to pursue it. In order to solve these two problems, a single platform should be available which will provide a perfect career guidance for the student community.

The is to figure out how to choose, identify, and explore their interests while also getting help selecting potential career choices and skill-building alternatives. Successful individuals in their field of interest will provide them with exciting advice in direct interaction with them. CareEx is an AI/ML-based help system that offers a Career dendrogram with many alternatives, including the real workforce requirements, broken down by category, for the public and private sectors. Our solution's primary purpose is to set professional objectives for users and estimate their likelihood of enrolling in the schools and institutions of their choice. With this guidance, many individuals will be able to see the future options that are open to them in the current situation as well as learn more about themselves, their strengths, and the ideal career path for them.

Keywords—Career Guidance, Eligibility Predictor, Data Prediction, Machine Learning

I. INTRODUCTION

Choosing the right career is crucial for an individual's future, but many aspirants struggle with self-realization and insufficient guidance, potentially leading to unsuccessful professional lives. Artificial Intelligence has made significant advancements globally and can play a pivotal role in addressing these challenges. AI models can provide virtual career guidance and help assess university acceptance rates, which is particularly valuable for foreign universities. However, such resources are often lacking in local education systems.

India, as one of the world's youngest nations, must prioritize these facilities to empower its aspirants and contribute to the nation's development. Personalized career guidance can be achieved by evaluating individual abilities and interests, creating a profile to identify strengths and weaknesses, and predicting suitable career paths based on this data.

A University Eligibility Predictor can also be developed by analyzing student data and university rankings, considering additional exams like NEET for medical admissions. Machine learning algorithms can be employed to train models using this data, as obtaining a larger dataset manually may be necessary for higher accuracy. By comparing various models, the most efficient algorithm can be selected for the proposed solution.

II. LITERATURE SURVEY

Students are given career counseling through a variety of methods. This section provides a thorough literature review of the numerous works that have been conducted.

A. CAREER GUIDANCE

In [1], the authors have analysed the career guidance in depth. The proposed system, the authors have analysed the skill sets of the students. Marks are allocated to the answers given by them by validating the answers, thereby predicting the suitable department, as per the score of the candidate's skill set. These are processed through KNN model. It provides overall result analysis of the performance of the candidate thus, the candidate can recognise where they are lacking sufficient skills, such that the candidate can improve those skills. In [2], the authors have proposed a system, where the skill sets of the students are analysed. The skill set of the candidates are validated through a serious of questions. These are processed through Random Forest model. It provides overall result analysis of the performance of the candidate thus, the candidate can recognise where they are lacking sufficient skills, such that the candidate can improve those skills. The proposed solution has a dataset that has been self-developed as there were not any values available. In [3], the authors have proposed a mobile based solution for career guidance. Students and parents can utilise the application since it offers comprehensive information on universities, including tuition and housing costs, eligibility requirements, campus employment opportunities, lodging options, scholarship programmes, campus support services, laws and regulations, etc. It aids students in choosing engineering and management-related universities in India and overseas. The proposed system uses database to store the information of the user. For these details provided by the user, the system filters, one which may narrow down the list of colleges based on the different requirements. In [4], the

authors have proposed a web-based career guidance system, which provides the prediction of the suitable path of the student through analyzing the major five skill sets. These are processed through the Random Forest model. It provides overall result analysis of the performance of the candidate. These are done by using the prediction module. Then, the system uses Microsoft Azure to develop, test and deploy the application. In [5], the authors proposed a solution considering the conscious and subconscious factors of the individual. It includes, Aptitude Test, Psychometric test, and handwriting analysis. To analyze the patterns and physical characteristics of the handwriting, Graphology is used. It is a system that implements integrated model of 3 different methods. In [6] the authors have proposed a web application that is used to guide the students and also helps the job seekers. The students undergo few mock tests, and the values are added to the database. According to the details available in the database, guidance is provided. In [7] the authors have proposed an application where the students can ask their questions related to the career, which will be sent to the appropriate professional to answer them. It is built using neural network and LightFM Hybrid Recommendation Model. In [8] the authors have proposed a web application where, the students go through levels of examinations with set of questions, in order to get their interest and suitable domains. The questions are usually related to 4 main domains.

B. UNIVERSITY ELIGIBILITY PREDICTOR

In [9] the authors have proposed an application where, the university admission is predicted using a Stacked Ensemble Model with 10 different machine learning models(layers). This includes various models ensembled as one and the model is trained along with the dataset. The dataset only contains the scores of major and famous exams like TOEFL, IELTS, GATE, etc., In [10] the authors have tested the dataset with four different models namely Naïve Bayes, Logistic Regression, KNN, Decision Tree. The results have been compared with each other to get the model which van give the highest accuracy. Here the dataset used are imported from a particular university enrolment.

III. PROPOSED SYSTEM

One of the categorization methods used in artificial intelligence is the decision tree. It aids in categorising the provided dataset and producing a prediction model. As a result, it can forecast the value of unobserved data. Decision trees categorise instances by arranging them in a tree from the root to a leaf node, which gives the instance's categorization. As seen in the above diagram, to classify an instance, one tests the attribute provided by the tree's root node before moving on to the branch that represents the attribute's value. The subtree rooted at the new node is then subjected to the same procedure once more.

The proposed system aims for providing an easier way for the students to know their career path, and also the probability of admission into their desired colleges/universities. It has two modules namely. Career Guidance and University Eligibility Prediction.

A. Dataset Collection

In the proposed system, there are two modules that are to be discussed. In Career Guidance module, the dataset is acquired from the Kaggle in order to provide the user i.e., the students with right career options. It has a set of questions, which are answered earlier by experts in that particular field. For now, the main two streams discussed are MEDICINE and ENGINEERING. Apart from these too, there are guidelines for other fields too like, EMPLOYEMENT, LAWYER, etc. For the University Eligibility Prediction module, as there are no datasets available for this, the dataset was self-created.

I. *Career Guidance*

In this section, let us discuss the process of acquiring of datasets for the mentioned module. We also see the how the datasets are classified into various fields.

Question title	Expert Answer
How long does it take to truly pursue a career in Psychology?	I think it really depends on how far you want to go with it and the state regulations where you live. Many require a Master's Degree in order to pursue you own practice, but I would start there. It also really depends on the career path you want to go. Psychology is a very open field to get into and it is within the Master's level where you tend to specialize more in an area! Good luck!
I'm interested in media but not sure what degrees or jobs I can get?	You must opt for Bachelors programs in Film, TV & Production. You can get jobs in print and electronic media both. In their production or content development department as well.
What is the hardest thing about becoming a lawyer?	The lawyer's response is, "it depends!" Each person is different. I thought the prep was okay, but I struggled with the legal writing and research. The entire process is designed to be difficult, but doable, if you are capable. Your best bet is to develop good study habits now, practice for the tests, draft your entry essay early, and once you're in, ask for help early and often, develop a network of peers to assist as you study, and be ready to work hard and read a lot!

Table 1: Dataset Gathering

II. Eligibility Predictor

Let's talk about how the dataset for the aforementioned module was created in this part. The supplied datasets mostly focus on international universities for the university eligibility predictor. The dataset is created using the SOA, GATE, TOEFL, and other scores in order to forecast the likelihood of admission. In contrast, the proposed approach places a greater emphasis on the regional institutions and their admissions standards. Cut-off scores from the board exam and the community that the applicant belongs to are used as deciding criteria for admission to tamilnadu colleges. There are currently no pre-available datasets for this. To create this, we gathered information on the colleges, branches that were accessible at those colleges, community divisions, and minimum and maximum cut-offs for admission. Values from about 6 colleges in Tamil Nadu are included in the raw data. A manually created dataset was created using the obtained data. This dataset functions effectively since it includes roughly 76,000 rows. The sample dataset is described in Tab 2:

COLLEGE_NAME	BRANCH_NAME	COMMUNITY	CUT-OFF	CHANCE OF ADMIT
SRI SIVASUBRAMANIYA NADAR COLLEGE OF ENGINEERING	BIO MEDICAL	OC	168.65	84.33
PSG COLLEGE OF ENGINEERING	CS	BC	197.58	98.79
SRI KRISHNA COLLEGE OF ENGINEERING	ECE	BC	181.36	90.68
THIAGARAJAR COLLEGE OF ENGINEERING	CS	MBC	175.58	87.79
KUMARAGURU COLLEGE	BIO TECH	OC	173.79	86.90
OF TECHNOLOGY				

Table 2: Sample dataset

I. Pre-Processing

All the collected dataset has some string values which cannot be used while training the machine learning model. All the columns should be of either float or int values. To achieve this, pre-processing has been done using LabelEncoder which is present in the sklearn of a machine learning model. Using this, the string values can be encoded and used to train the model. The encoded dataset is as described in Tab 3:

COLLEGE_NAME	BRANCH_NAME	COMMUNITY	CUT-OFF	CHANCE OF ADMIT
1	1	1	168.65	84.33
2	2	2	197.58	98.79
3	4	2	181.36	90.68

Table 3: Encoded dataset Sample

IV. RESULT DISCUSSIONS AND OUTCOMES

The specifics and outcomes of the suggested machine learning with the three machine learning models are detailed in this section. Using the methods for Decision Tree, Random Forest, and KNN, the suggested system is validated. They are carried out on a workstation with 16GB of Memory and an Intel Core i7 processor from the 12th generation. The dataset was gathered, and then pre-processed in accordance with the necessary criteria. By creating the dataset, the raw data was expanded, yielding a dataset with 76,000 rows.

A. Performance Metrics

Assessing the proposed model is the most crucial stage in developing an effective deep learning model. Important classification metrics including Precision, Accuracy, F1-score, and Recall have been calculated in order to achieve this. This computation is based on the sums of the True Positive, False Positive, True Negative, and False Negative values. These are the equations used to calculate these metrics.

Accuracy is the total percentage reached for the model's reliable prediction. This measurement, which is the most important of all, determines the model's effectiveness. The formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

A Recall is a sensitivity rate, which shows how many predictions it has accurately got out of all of them.

$$Recall = \frac{TP}{TP + FN}$$

Prediction is the proportion of correctly identified samples, or "True Positives," to all of the samples that have been categorized as positive.

$$Precision = \frac{TP}{TP + FP}$$

The harmonic mean of the proposed model's recall and accuracy is known as the F1-score. The recommended model performs better when categorization metrics increase.

$$F1 - score = \frac{2TP}{2TP + FP + FN}$$

B. Experiment 1: Model Assessment Using a Decision Tree

The generated dataset is trained with the Decision tree method in the first experiment. In this case, the pre-processed dataset is used as input for the model. Table 4 displays the experiment's outcomes. The table shows that the model has an overall accuracy of 88%.

Label	Precision	Recall	F1-Score	Support
1	0.80	0.83	0.78	55
2	0.67	0.87	0.79	67
3	0.80	1.00	0.84	67
4	0.83	0.85	0.88	43
Accuracy		0.88		30400
Macro avg.	0.91	0.88	0.90	30400
Weighted avg.	0.88	0.88	0.88	30400

Table 4: Result of system with Decision Tree

C. Experiment 2: Model Assessment Using Random Forest

The resulting dataset is trained using the Random Forest method in the second experiment. In this case, the pre-processed dataset is used as input for the model. Table 5 summarizes the experiment's findings. The table shows that the model has an overall accuracy of 79%.

Label	Precision	Recall	F1-Score	Support
1	0.76	0.70	0.82	34
2	0.85	0.63	0.72	27
3	0.92	1.00	0.71	48
4	0.80	0.75	0.77	43
Accuracy		0.79		30400
Macro avg.	0.82	0.82	0.82	30400
Weighted avg.	0.80	0.79	0.79	30400

Table 5: Result of system with Random Forest

D. Experiment 3: Model Assessment Using KNN

The created dataset is trained using the KNN method in the third experiment. In this case, the pre-processed dataset is used as input for the model. The experiment's findings are shown in Table 6. The table shows that the model has an overall accuracy of 65%.

Label	Precision	Recall	F1-Score	Support
1	0.57	0.72	0.79	38
2	0.63	0.67	0.70	42
3	0.50	1.00	0.65	34
4	0.68	0.72	0.77	46
Accuracy		0.65		30400
Macro avg.	0.70	0.72	0.72	30400
Weighted avg.	0.66	0.65	0.65	30400

Table 6: Result of system with KNN

Each model's variances in accuracy are the result of various improvements. Different metrics are calculated for each model. Three categorization models—Decision Tree, Random Forest, and KNN—are used by the proposed system. In this case, three distinct classifications result in

three distinct outputs. A single tree with roots and leaves is provided by Decision Tree. Whereas the KNN classifier has class membership and the Random Forest technique consists of several decision trees. As a result, each of the three models has a distinct accuracy rate. Also, the measurements adjust according on the dataset. The comparison of the metrics in all the three models are represented in the Figure. 2 below.
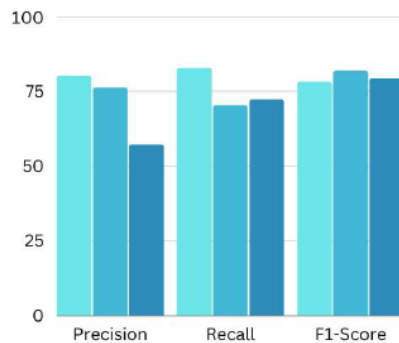

Figure 2: Comparative study of metrics: precision, recall, and F1-score.

E. Experiment 4: Comparative Evaluation of Various Machine Learning Algorithms

A comparison of a few machine learning models, such as Decision Tree, Random Forest, and KNN, has been applied over the suggested system. The statistic displays the accuracy of the suggested model by comparing the number of true positives to the number of true positives and false negatives. It also indicates the accuracy of the prediction model. As seen in Fig.3, the KNN model has the lowest accuracy, while the Decision tree has the highest.

Figure 3: Comparison of accuracy for Decision tree, Random Forest and KNN models

V. CONCLUSION

The proposed work is an application for the young minds to provide career guidance. They will know the details of the options available in all locations. They can choose the best and the one that suits them. Our solution's major purpose is to provide professional objectives and opportunities, as well

as a probability forecast for enrolment in their selected colleges/universities. This assistance can help many people to visualize future paths available on this current scenario and to know more about what they dream to become, and the path towards achieving the same. After undergoing various experiments, the results indicate that the suggested method has an accuracy of 88%, which is higher than any other system now in use. Furthermore, because the suggested system is self-generated, it has a unique dataset. They distinguish the proposed system from the existing ones.

In the future, the career guidance module can be made more interactive, where it could be trained with an appropriate machine learning model, and make students or the user search and get help using few keywords. Also, more no of colleges and branches are to be added to make the proposed system more efficient and works best for all category of students. The system is to be enhanced with respect to other fields like Medicine, Arts, and many more.

REFERENCES

[1] S. Vignesh, C. Shivani Priyanka, H. Shree Manju and K. Mythili, "1," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 987-990,doi:10.1109/ICACCS51430.2021.9441978.

[2] A. Kamal, B. Naushad, H. Rafiq and S. Tahzeeb, "2," 2021 4th International Conference on Computing & Information Sciences (ICCIS), Karachi, Pakistan, 2021, pp. 1-7, doi: 10.1109/ICCIS54243.2021.9676408.

[3] V. D. Kamath, A. Meher, V. Vidhya and S. Deepthi, "An Android based Mobile Application for Career Guidance," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 854-860, doi: 10.1109/ICICCT.2018.8473128.

[4] Anand, D. Patil, S. Bhaawat, S. Karanje and V. Mangalvedhekar, "Automated Career Guidance Using Graphology, Aptitude Test and Personality Test," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697642.

[5] K. Seng and A. M. Zeki, "Career Guidance and Employment Management System," 2014 3rd International Conference on Advanced Computer Science Applications and Technologies, Amman, Jordan, 2014, pp. 73-78, doi: 10.1109/ACSAT.2014.20.

[6] I. Dutta, A. Shah, M. M. Safa and K. Jayavel, "Building a Recommendation System for Career Advice for Students from Professionals," 2022 International Conference on Computer Communication and

Informatics (ICCCI), Coimbatore, India, 2022, pp. 1-10, doi: 10.1109/ICCCI54379.2022.9740837.

[7] K. F. Mohammed, S. Gawade and V. Nimkar, "Online career counseling system using adaptive e-learning," 2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE), Bhopal, India, 2017, pp. 583-592, doi: 10.1109/RISE.2017.8378222.

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

[9] B. Ujkani, D. Minkovska and L. Stoyanova, "A Machine Learning Approach for Predicting Student Enrollment in the University," 2021 XXX International Scientific Conference Electronics (ET), Sozopol, Bulgaria, 2021, pp. 1-4, doi: 10.1109/ET52713.2021.9579795.

[10] A. Sivasangari, V. Shivani, Y. Bindhu, D. Deepa and R. Vignesh, "Prediction Probability of Getting an Admission into a University using Machine Learning," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1706-1709, doi: 10.1109/ICCMC51019.2021.9418279.