Dr. Shabeen Taj G. A[1*], Midthur A Salman Khan[2*], Dr. Waseeha Firdose[3*]

An Integrated Deep Learning Framework for Personalized Career Guidance and Course Recommendation

Journal of Electrical Systems

Abstract — The transition from academia to a professional career is a critical phase for students, often marked by uncertainty in selecting an appropriate job role. This uncertainty is compounded by a lack of awareness of the requisite skills and the vast array of available learning resources. Many organizations and corporate entities offer high-quality, free courses on platforms like NPTEL, Coursera, edX, Microsoft Learn, and Google Learning, yet students frequently remain unaware of these opportunities. This paper proposes an integrated system designed to mitigate these challenges. The system employs a Deep Neural Network (DNN) model to predict suitable Information Technology (IT) job roles for students based on their academic profiles, skills, and interests, achieving a prediction accuracy of 90.87%. Subsequently, a content-based course recommendation engine suggests relevant, free courses tailored to the predicted job role, leveraging cosine similarity for optimal matching. By consolidating courses from numerous free platforms into a single dataset, the system provides a centralized hub for skill development. This approach not only offers personalized career guidance but also facilitates access to valuable educational content, thereby bridging the gap between academic knowledge and industry requirements and empowering students to make informed career decisions.

Keywords— *Job Role Prediction, Course Recommendation, Deep Neural Network (DNN), Content-Based Filtering, Cosine Similarity, Educational Data Mining.*

I. Introduction

The culmination of a student's academic journey often leads to a pivotal question: "What next?" This dilemma is particularly pronounced in vast and dynamic fields like engineering and information technology, where the spectrum of potential career paths is wide and continuously evolving [1], [2]. A significant number of graduates struggle to align their academic experiences with the specific demands and roles of the industry, leading to potential skill mismatches and career dissatisfaction [3]. This underscores the critical need for intelligent systems that can provide data-driven career guidance.

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative technologies, revolutionizing sectors such as healthcare, finance, and notably, education [4], [5]. Within education, AI applications range from learning analytics and automated grading to personalized tutoring systems [6]. Recommender systems, a cornerstone of AI, are powerful tools that analyze patterns in user data to suggest relevant items, such as products, movies, or, in this context, educational content and career paths [7], [8].

While several platforms offer career assessments, many lack a strong predictive analytical foundation or fail to integrate seamless access to subsequent learning resources [9], [10]. Furthermore, although an abundance of free, high-quality courses exists on platforms offered by universities (e.g., MIT OpenCourseWare, NPTEL) and corporations (e.g., Microsoft Learn, Google Learning), students often face a discovery problem [11]. These valuable resources are siloed across different websites, and commercial platforms often prioritize promoting paid content, leaving free courses underutilized.

[1]Assistant Professor, Department of Computer Science and Engineering,Government Engineering College,Ramanagara, Karnataka, India, 562159
Email ID : shab2en@gmail.com
[2]Assistant professor, Department of Mechanical engineering, Sanskrit school of engineering
Email ID : m.a.salman087@gmail.com
[3]Assistant professor, Government Ramnayayan Chellaram college of Commerce and Management Bengaluru
Email ID : waseeha.ha@gmail.com

This paper addresses these gaps by presenting a unified system with two core components:

1. **A Job Role Prediction Engine:** Utilizing a Deep Neural Network (DNN) to analyze a student's profile and accurately predict the most suitable IT job role from a set of 34 possibilities.

2. **A Course Recommendation Engine:** Employing content-based filtering techniques to recommend a curated list of free courses specifically tailored to bridge the skill gap for the predicted job role.

The primary contribution of this work is the integration of a highly accurate predictive model with a practical recommendation system, creating an end-to-end solution that guides students from career ambiguity to a clear upskilling path using freely available educational resources.

The rest of this paper is organized as follows: Section II discusses the related work. Section III describes the datasets used. Section IV details the proposed methodology and algorithm. Section V presents the results and discussion. Finally, Section VI concludes the paper and suggests future directions.

II. Related Work

The application of data mining and machine learning in education, often termed Educational Data Mining (EDM) and Learning Analytics (LA), has gained substantial traction. Previous research has explored various methods for predicting student performance and outcomes. For instance, [12] used Support Vector Machines (SVM) to predict student performance, while [13] applied models to predict career choices based on campus big data, analysing student behaviour patterns.

In the realm of career prediction, research has often focused on psychological assessments or academic grades. [14] proposed a model based on Adversity Quotient and Career Interest, and [15] used basic data mining techniques for career prediction. However, these approaches often lack the depth of modern deep learning techniques for handling complex, multi-feature datasets specific to IT roles.

Recent advancements have seen the application of deep learning for more nuanced predictions. [16] explored deep neural network models for job prediction applications, demonstrating the superiority of DNNs over traditional models in capturing non-linear relationships between skills and job titles.

On the recommendation side, course recommendation systems have been built primarily using collaborative filtering, content-based filtering, or hybrid methods. [17] developed a skill-based course recommendation system, and [18] built a model for students using machine learning. Content-based approaches, which recommend items similar to those a user has liked in the past, are particularly effective in cold-start scenarios where user history is limited [19], [20]. The use of cosine similarity for text-based recommendation is a well-established technique in information retrieval and recommender systems [21], [22].

Our work distinguishes itself by combining a high-accuracy DNN prediction model for a wide array of IT job roles with a content-based recommender system that specifically aggregates and suggests free, high-quality courses, addressing both the career guidance and skill development needs of students in a single framework.

III. Dataset

The system leverages two distinct datasets, one for each core component.

A. Job Role Prediction Dataset

The dataset for IT job role prediction was sourced from Kaggle [23]. It comprises 20,000 distinct records, each with 36 features that are critical for prediction, such as technical skills, programming languages, soft skills, and educational background. The data was aggregated from various sources, including profiles of working professionals on LinkedIn and organizational surveys. The target variable is the job title, distributed across 34 unique IT roles. The distribution is highly balanced, as detailed in Table I, which was crucial for training an unbiased model. This balance prevents the model from being skewed towards more frequent job roles.

TABLE I. DISTRIBUTION OF IT JOB ROLES IN THE DATASET

Sl. No	IT Job Title	#Samples	Percentage (%)
1	Applications Developer	551	2.7550%
2	Business Intelligence Analyst	540	2.7000%
3	Business Systems Analyst	582	2.9100%
4	CRM Business Analyst	584	2.9200%
5	CRM Technical Developer	567	2.8350%
6	Data Architect	564	2.8200%
7	Database Administrator	593	2.9650%
8	Database Developer	581	2.9050%
9	Database Manager	570	2.8500%
10	Design & UX	588	2.9400%
11	E-Commerce Analyst	546	2.7300%
12	Information Security Analyst	543	2.7150%
13	Information Technology Auditor	558	2.7900%
14	Information Technology Manager	591	2.9550%
15	Mobile Applications Developer	538	2.6900%
16	Network Engineer	621	3.1050%
17	Network Security Administrator	1112	5.5600%
18	Network Security Engineer	630	3.1500%
19	Portal Administrator	593	2.9650%
20	Programmer Analyst	529	2.6450%
21	Project Manager	602	3.0100%
22	Quality Assurance Associate	565	2.8250%
23	Software Developer	587	2.9350%
24	Software Engineer	590	2.9500%
25	Software Quality Assurance (QA) / Testing	571	2.8550%
26	Software Systems Engineer	575	2.8750%
27	Solutions Architect	578	2.8900%
28	Systems Analyst	550	2.7500%
29	Systems Security Administrator	562	2.8100%
30	Technical Engineer	557	2.7850%
31	Technical Services/Help Desk/Tech Support	558	2.7900%
32	Technical Support	565	2.8250%
33	UX Designer	589	2.9450%
34	Web Developer	570	2.8500%
	Total	**20000**	**100.0000%**

B. Course Recommendation Dataset

For the recommendation engine, a novel dataset was curated by aggregating information on free courses from multiple reputable learning platforms, including:

- NPTEL
- MIT OpenCourseWare
- Coursera (free courses)
- edX (free audit track)
- Microsoft Learn

- Google Learning
- Future Skill Prime
- Infosys Springboard

The dataset includes key features such as Course_ID, Course_Title, Domain, URL, Course_Description, Difficulty_Level, Number_of_Lectures, Offering_Platform, Availability (free/paid), and Languages_Tools_Used. A sample of this dataset is shown in Fig. 1. This consolidation provides a comprehensive repository of free learning materials across various tech domains. This consolidation provides a comprehensive repository of free learning materials across various tech domains.

	A	B	C	D	E	F	G	H	I	J	K
1	Course_ID	Course_Title	Domain	URL	Course_Description	Difficulty	No.of_Lectures	Course_Duration	Course_Offered	Availability	Languages_and_To
2	1	Responsive Web Design	Web Development	https://www.freecc	In this Responsive W	Beginner, I	189	300 Hours	Free Code Cam	Free	HTML, HTML5,CS
3	2	JavaScript Algorithms and [	Web Development	https://www.freecc	While HTML and CS	Beginner, I	299	300 Hours	Free Code Cam	Free	JavaScript
4	3	Python for Beginners	Python Developer	https://www.solole	Python is a popular,	Beginner	41	80 Hours	Solo learn	Free	Python
5	4	Front End Development Libr	Web Development	https://www.freecc	Now that you're fami	Beginner, I	132	300 Hours	Free Code Cam	Free	Bootstrap, Rect, R
6	5	C++ basics	Game Developmen	https://www.solole	Our C++ tutorial cove	Beginner	84	200 Hours	Solo learn	Free	C++
7	6	Data Visualization	Data Analysis	https://www.freecc	Data is all around us	Beginner, I	39	300 Hours	Free Code Cam	Free	JSON, API,JavaSc
8	7	Back End Development and	Bankend Develome	https://www.freecc	Until this point, you'v	Beginner, I	34	300 Hours	Free Code Cam	Free	Node.js
9	8	Java Tutorial	Java Developer	https://www.solole	With our interactive J	Beginner	70	200 Hours	Solo learn	Free	Java
10	9	Quality Assurance	Web Development	https://www.freecc	As your programs or	Intermediat	47	300 Hours	Free Code Cam	Free	Node.js,Express,P
11	10	Scientific Computing with P	Software Developn	https://www.freecc	Python is one of the	Beginner, I	56	300 Hours	Free Code Cam	Free	Python
12	11	C# Tutorial	Software Developn	https://www.solole	With our interactive (	Beginner	80	300 Hours	Solo learn	Free	c#
13	12	Data Analysis with Python	Data Analysis	https://www.freecc	Data Analysis has be	Beginner, I	37	300 Hours	Free Code Cam	Free	Python,Numpy
14	13	Information Security	System Security A	https://www.freecc	With everything we d	Beginner	21	300 Hours	Free Code Cam	Free	HelmetJs,Python,N
15	14	Machine Learning with Pyth	Machine Learning I	https://www.freecc	Machine learning has	Beginner, I	36	300 Hours	Free Code Cam	Free	Python,Numpy,Ter
16	15	JavaScript Tutorial	Web Development	https://www.solole	Learn all the basic fe	Beginner	65	180 Hours	Solo learn	Free	JavaScript
17	16	Coding Interview Prep	Software Developn	https://www.freecc	If you're looking for f	Intermediat	716	1000 Hours	Free Code Cam	Free	JavaScript
18	17	Python Core	Python Developer	https://www.solole	Learn Python, one of	Beginner, I	102	300 Hours	Solo learn	Free	Python

AA5

Course Data Set - dataset

Fig. 1. Sample of the Course Recommendation Dataset

IV. Proposed Methodology and Algorithm

The overall architecture of the proposed system is depicted in Fig. 2. The process is segmented into three primary phases: Data Preparation, Job Role Prediction, and Course Recommendation.

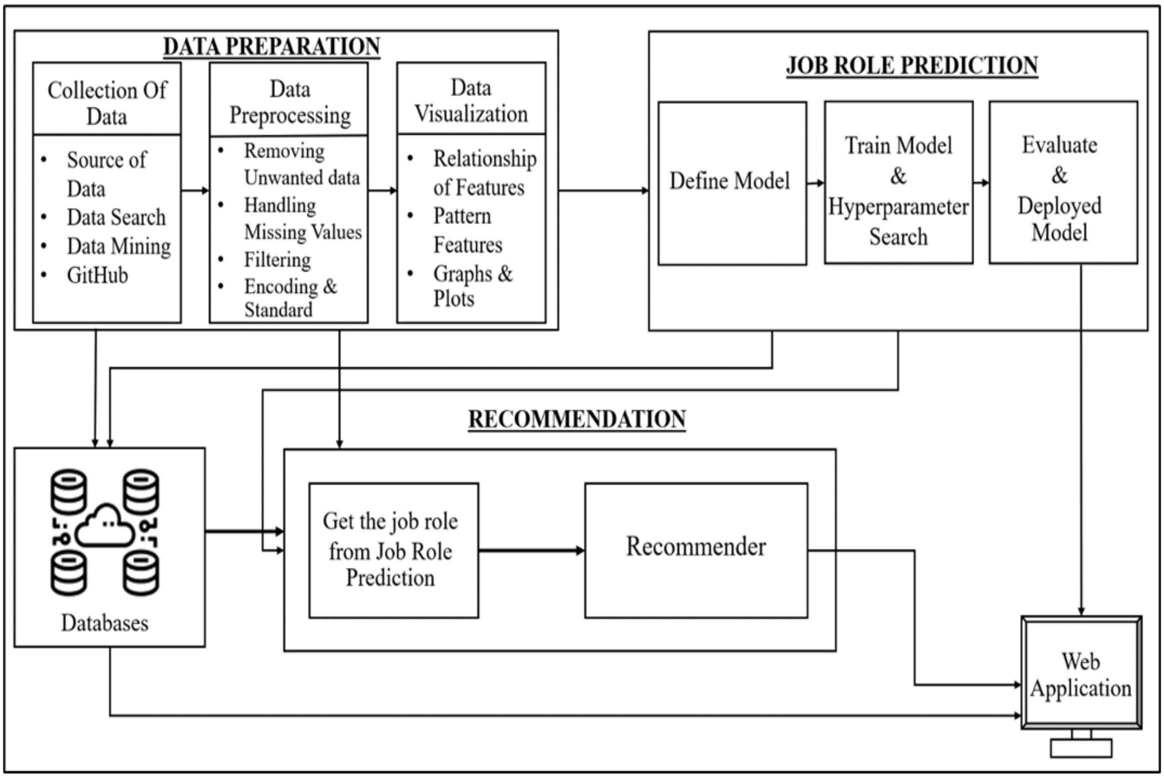

Fig. 2. High-Level Architecture of the Proposed System.

A. Data Preprocessing

The raw data from both datasets required significant preprocessing to be suitable for modelling. For the job prediction dataset, the following steps were taken:

1. **Encoding:** Categorical features were converted into numerical values using Label Encoding. The target variable (job role) was transformed into a binary matrix representation using One-Hot Encoding [24].

2. **Scaling:** All numerical features were standardized to have a mean of 0 and a standard deviation of 1 using StandardScaler [25] to ensure uniform contribution to the model training.

3. **Train-Test-Validation Split:** The preprocessed data was partitioned into 70% for training, 20% for testing, and 10% for validation.

The course dataset required text-specific preprocessing:

1. **Text Cleaning:** The Course_Description and Languages_Tools_Used fields were cleaned using the neattext library to remove special characters, stop words, and perform stemming.

2. **Feature Vectorization:** The cleaned text was converted into a numerical matrix using Count Vectorization, which creates a bag-of-words model.

B. Job Role Prediction using Deep Neural Network (DNN)

The proposed DNN architecture consists of:

- **Input Layer:** Accepts the 36-dimensional feature vector.

- **Hidden Layers:** Three fully connected (Dense) layers with 272, 136, and 68 neurons, respectively. These numbers were chosen following a power-of-two scaling convention and optimized through hyperparameter tuning.

- **Output Layer:** A dense layer with 34 neurons (one for each job role) using the Softmax activation function to output a probability distribution over all possible roles.

The Glorot Uniform initializer [26] was used to initialize weights. The Rectified Linear Unit (ReLU) [27] was chosen as the activation function for all hidden layers due to its computational efficiency and effectiveness in mitigating the vanishing gradient problem. The model was compiled with the Adam optimizer [28] (with a learning rate of 0.001) and categorical cross-entropy as the loss function.

Hyperparameter tuning was performed using the Talos library [29] to optimize batch size, number of epochs, and regularization parameters. The final model was trained with a batch size of 32 for 118 epochs.

Algorithm 1: Job Role Prediction Algorithm
Input: Student Feature Vector *X*
Output: Predicted Job Role *Y*
1: Preprocess input vector X (Label Encoding + StandardScaler)
2: Load trained DNN model *M*
3: probabilities=M.predict(*X*)

4: *index* =argmax(probabilities)

5: *Y*=*decode_onehot*(*index*) // Map index to job role label
6: **return** *Y*

C. Course Recommendation using Content-Based Filtering

For a predicted job role *Y*, the system queries the course database to find all courses tagged with domain(s) relevant to *Y*. The core of the recommendation is a cosine similarity metric. The system creates a combined text corpus for each course from its title, description, and tools. This corpus is vectorized into a term-frequency matrix *C*.

The similarity between courses is calculated using cosine similarity, which measures the cosine of the angle between two vectors in a multi-dimensional space. Given two course vectors $\vec{V}$ and $\vec{\boldsymbol{W}}$, the similarity is calculated as:

$$Cosine(\vec{v},\vec{w}) = \frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\sum_{i=1}^{N} v_i \text{X} w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

A subset of "seed courses" known to be highly relevant to job role *Y* is selected. The system then computes the cosine similarity between every other course and these seed courses. The top *N* courses with the highest average similarity scores are recommended to the user.

Algorithm 2: Course Recommendation Algorithm
Input: Predicted Job Role *Y*, Course Dataset *D*
Output: List of Recommended Courses *R*
1: *seed_courses*=*get_seed_courses*(*Y*)
2: *corpus*=*create_combined_text_corpus*(*D*)
3: *vectorizer*=*CountVectorizer*().*fit*(*corpus*)
4: *matrix*=*vectorizer*.*transform*(*corpus*)
5: *sim_scores*=[]
6: **for** each course *c* in *D* **do**
7: *avg_sim*=*average*(*cosine_similarity*(*c*,*sc*) forall *sc* in *seed_courses*)
8: *append*(*avg_sim*,*c*) to *sim_scores*
9: **end for**
10: *sort*(*sim_scores*) // sort by average similarity score descending
11: *R*=*top_N*(*sim_scores*) // get top N courses
12: **return** *R*

V. Results and Discussion

The DNN model for job role prediction was trained and evaluated rigorously. The training and validation accuracy and loss curves are shown in Fig. 3. The curves show a healthy learning process: both training and validation accuracy increase steadily while loss decreases, with no significant gap between them, indicating that the model generalizes well without overfitting.

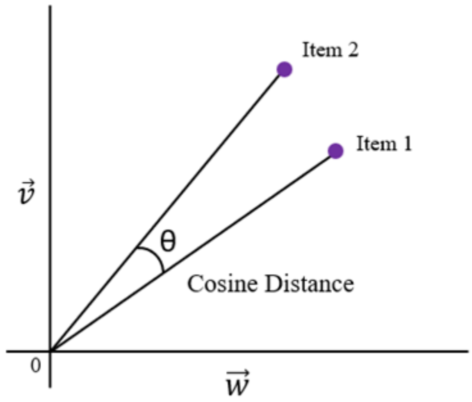

Fig. 3. Model Accuracy (Left) and Loss (Right) curves during training.

The model achieved a final test accuracy of **90.87%**, demonstrating its effectiveness in classifying students into the correct IT job role based on their features. To further analyze the model's performance across all classes, a normalized confusion matrix for the test set predictions is presented in Fig. 4. The strong diagonal indicates high per-class accuracy. Minor misclassifications occur between semantically similar roles (e.g., between various database administrators or software developers/engineers), which is an expected and common challenge in multi-class classification [30]. This high level of accuracy provides a reliable foundation for the subsequent course recommendation step.

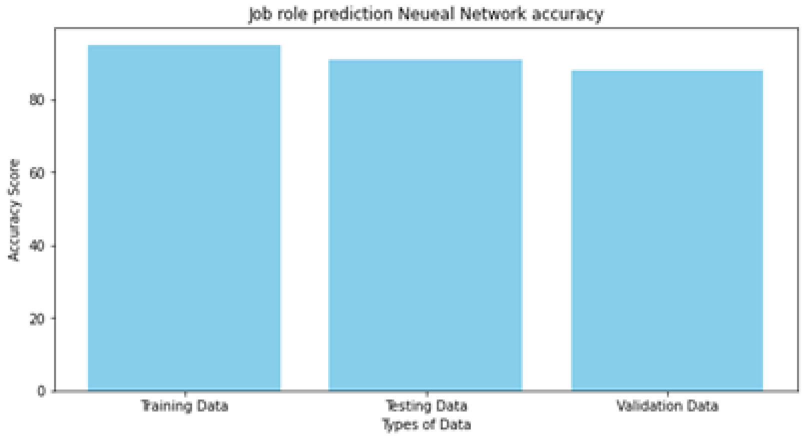

Fig. 4. Normalized Confusion Matrix for Job Role Prediction on the Test Set.

The course recommendation engine was tested qualitatively. For a predicted job role (e.g., "Data Scientist"), the system successfully recommended relevant free courses on Python, Statistics, Machine Learning, and specific tools like TensorFlow from platforms like Coursera, edX, and Microsoft Learn. The use of cosine similarity ensured that the recommended courses were contextually aligned with the required skills for the job role.

VI. Conclusion and Future Work

This paper presented an integrated system that leverages deep learning and content-based filtering to address two critical challenges faced by students: career choice ambiguity and access to relevant learning resources. The developed DNN model predicts IT job roles with high accuracy (90.87%), and the recommendation engine effectively maps these roles to freely available online courses.

The key contribution of this work is the creation of a seamless pipeline from career prediction to skill development. By aggregating free courses from disparate sources, the system acts as a valuable centralized resource, promoting accessibility and continuous learning.

For future work, the system can be enhanced in several ways:

1. **Incorporating Collaborative Filtering:** Integrating user feedback (likes, clicks, completions) could lead to a hybrid recommendation model, improving personalization over time.

2. **Real-Time Data Integration:** Implementing web scrapers to dynamically update the course database would ensure the recommendations remain current with the latest course offerings.

3. **Skill Gap Analysis:** A more advanced feature could involve analyzing a student's current skill profile against the requirements of a predicted job role to recommend courses that specifically target identified skill gaps.

4. **Expanding Beyond IT:** The methodology can be extended to other disciplines such as business, finance, or healthcare, making the system versatile for a wider student population.

References

[1] M. Y. Arafath, M. Saifuzzaman, S. Ahmed and S. A. Hossain, "Predicting Career Using Data Mining," in *2018 International Conference on Innovation in Engineering and Technology(ICIET)*, 2018.

[2] R. H. Rangnekar, K. P. Suratwala, S. Krishna and S. N. Dhage, "Career Prediction Model Using Data Mining and Linear Classification," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018.

[3] K. B. Eashwar, R. Venkatesan and D. Ganesh, "Student Performance Prediction Using SVM," *International Journal of Mechanical Engineering and Technology (IJMET)*, vol. 8, no. 11, pp. 649–662, 2017.

[4] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access*, vol. 8, pp. 75264-75278, 2020.

[5] J. Estevez, G. Garate, and M. Graña, "Gentle Introduction to Artificial Intelligence for High-School Students Using Scratch," *IEEE Access*, vol. 7, pp. 179027-179036, 2019.

[6] B. R. Barber and M. Sharkey, "Course Correction: Using Analytics to Predict Course Success," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 2012.

[7] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, 2nd ed. Springer, 2015.

[8] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*. Cambridge University Press, 2010.

[9] M. Nie, Z. Xiong, R. Zhong, W. Deng and G. Yang, "Career Choice Prediction Based on Campus Big Data," *Journal of Applied Science*, pp. 1-14, 2020.

[10] M. F. Lee, N. F. Mat Nawi and C. S. Lai, "Engineering student Job Performance Prediction Model based on Adversity Quotient and Career Interest," in *2017 IEEE 8th Control and System Graduate Research Colloquium (ICSGRC)*, 2017.

[11] V. Sankhe, J. Shah, T. Paranjape and R. Shankarmani, "Skill Based Course Recommendation System," in *2020 IEEE Bombay Section Signature Conference (IBSSC)*, 2020.

[12] K. B. Eashwar, R. Venkatesan and D. Ganesh, "Student Performance Prediction Using SVM," *International Journal of Mechanical Engineering and Technology (IJMET)*, vol. 8, no. 11, 2017.

[13] M. Nie, Z. Xiong, R. Zhong, W. Deng and G. Yang, "Career Choice Prediction Based on Campus Big Data," *Journal of Applied Science*, 2020.

[14] M. F. Lee, N. F. Mat Nawi and C. S. Lai, "Engineering student Job Performance Prediction Model based on Adversity Quotient and Career Interest," in *2017 IEEE 8th Control and System Graduate Research Colloquium (ICSGRC)*, 2017.

[15] M. Y. Arafath, M. Saifuzzaman, S. Ahmed and S. A. Hossain, "Predicting Career Using Data Mining," in *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, 2018.

[16] T. V. Huynh, K. V. Nguyen, N. L.-T. Nguyen and G.-T. Nguyen, "Job Prediction: From Deep Neural Network Models to Applications," in *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2020.

[17] V. Sankhe, J. Shah, T. Paranjape and R. Shankarmani, "Skill Based Course Recommendation System," in *2020 IEEE Bombay Section Signature Conference (IBSSC)*, 2020.
[18] L. P. J. Rani, D. C. J. W. Wise, K. V. Ajayram, T. Gokul and B. Kirubakaran, "Course Recommendation for student using Machine learning," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020.
[19] M. J. Pazzani and D. Billsus, "Content-Based Recommendation Systems," in *The Adaptive Web*, P. Brusilovsky, A. Kobsa, W. Nejdl, Eds. Springer, 2007.
[20] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based Recommender Systems: State of the Art and Trends," in *Recommender Systems Handbook*, F. Ricci et al., Eds. Springer, 2011.
[21] B. Li and L. Han, "Distance Weighted Cosine Similarity Measure for Text Classification," in *International Conference on Intelligent Data Engineering and Automated Learning (IDEAL)*, 2013.
[22] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," in *Proceedings of the First Instructional Conference on Machine Learning*, 2003.
[23] "Career Dataset for IT Jobs," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/davincisdata/careerdataset.
[24] K. Potdar, T. S. Pardawala, and C. D. Pai, "A comparative study of categorical variable encoding techniques for neural network classifiers," *Int. J. Comput. Appl.*, vol. 175, no. 4, pp. 7-9, 2017.
[25] Scikit-learn: Machine Learning in Python, Pedregosa et al., *JMLR*, vol. 12, pp. 2825-2830, 2011.
[26] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 2010.
[27] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 2011.
[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[29] Talos, "Hyperparameter Optimization for Keras/TensorFlow," [Online]. Available: https://github.com/autonomio/talos.
[30] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168-192, 2020.