The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

Author Name	Nidhi Mannopantar
Title	AI Based Career Analyzing Model
Paper/Submission ID	3594324
Submitted by	charankumarba@gmail.com
Submission Date	2025-05-08 13:43:05
Total Pages, Total Words	7, 2751
Document type	Research Paper

Result Information

Similarity **0 %**

1 10 20 30 40 50 60 70 80 90

Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File

DrillBit Similarity Report

0	0	A	A-Satisfactory (0-10%)
SIMILARITY %	MATCHED SOURCES	GRADE	B-Upgrade (11-40%)
			C-Poor (41-60%)
			D-Unacceptable (61-100%)

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE

AI Based Career Analyzing Model

Nidhi Mannopantar , Pooja M

Student, Department of Computer Science and Engineering, The Oxford College Of Engineering, Bangalore-560068, Karnataka, India

nidhicse2026@gmail.com ,poojamurthycse2026@gmail.com

Abstract - This paper explores the intersection of career goals, skill sets, and salary expectations among individuals using machine learning and data analytics techniques. By analyzing a dataset that includes personal profiles, skills, and expected salaries, the study identifies patterns and provides predictive insights. Visualizations such as heatmaps, bar charts, and box plots were used to reveal trends in salary distribution, skill demand, and field-wise growth opportunities. The results aim to assist job seekers, academic institutions, and industry stakeholders in making data-driven decisions regarding career planning and talent acquisition.

Keywords- Artificial Intelligence, Machine Learning, Career Recommendation System, Career Prediction, Personalized Career Counseling, Predictive Analysis

I. Introduction

Choosing the right career has never been easy—and in today's fast-changing world, it's more overwhelming than ever. With so many new job roles, industries, and technologies emerging, it's completely normal for students and even working professionals to feel confused or unsure about which path to follow.

That's where an **AI-Based Career Analyzing Model** can make a real difference. Imagine having a smart, personalized career advisor that understands your strengths, interests, educational background, and even your personality. Using artificial intelligence and machine learning, this model takes all of that information and turns it into meaningful career suggestions tailored just for you.

It doesn't stop there. The system can also highlight areas where you might need to improve and suggest relevant courses or training to help you grow. In short, it acts like a mentor—guiding you toward a career that fits not just the job market, but also who you are.

By helping people make more informed and confident career choices, this AI-powered tool can lead to more fulfilling and successful professional journeys.

II. Literature Review

In recent years, machine learning has played a big role in helping us understand employment trends—especially when it comes to predicting which jobs might suit someone best or estimating how much a person could earn in a certain role. Many studies have used classification models to match people with jobs based on things like their education, skills, and work experience. Others have used regression techniques to forecast salaries by looking at patterns in industry data. However, most of these studies tend to focus on specific sectors or job types, and often leave out important personal factors. For example, they might consider someone's technical skills but ignore their soft skills or long-term career goals. As a result, these models can miss the full picture of what makes a job the "right fit."

This paper takes a step further by looking at job prediction more holistically. Instead of just focusing on technical qualifications, it also considers soft skills—like communication and teamwork—and individual career aspirations. By combining these different aspects, our approach aims to offer more personalized and meaningful insights, helping people find roles that not only match their abilities but also align with where they want to go in their careers.

III. Dataset Overview and Preprocessing

Dataset Overview: The dataset used in this study is comprehensive and contains several critical fields that are vital for analyzing career trajectories and salary expectations. These fields include Name, Age, Degree, Field of Study, Years of Experience, Current Role, Career Goal, Expected Salary, and a set of both technical and soft skills. Each of these fields plays a pivotal role in understanding an individual's career profile and its relationship to salary predictions and career domain classification. For example, fields like Age, Years of Experience, and Current Role provide quantitative insights into career progression, while Degree and Field of Study offer valuable context about an individual's educational background, which may influence their earning potential.

Data Preprocessing: Before analyzing the dataset, several essential data preprocessing steps were performed to ensure the data was clean, consistent, and ready for analysis. One of the first challenges was handling missing values. Incomplete records, particularly in fields like Years of Experience or Field of Study, were either imputed using statistical methods (e.g., median imputation) or removed if they were too sparse to provide reliable information.

Next, the technical and soft skills listed in the dataset were normalized to a binary format for uniformity. This process involved converting each skill into a binary variable (1 if the skill was present, 0 if absent). For example, if a user had experience with Python, the dataset would reflect this as a 1 under the "Python" column; otherwise, it would be a 0. This transformation allowed the model to handle skill data consistently and enabled us to quantitatively assess the presence or absence of each skill across the dataset.

In addition, categorical data—such as Degree, Field of Study, and Current Role—was encoded into numerical values. Techniques such as One-Hot Encoding or Label Encoding were applied to transform non-.

Exploratory Data Analysis (EDA): With the data preprocessed, Exploratory Data Analysis (EDA) was performed to gain a deeper understanding of the relationships and patterns within the dataset. EDA involved generating various visualizations such as histograms, scatter plots, and box plots to explore the distribution of key variables like Age, Years of Experience, and Expected Salary. Correlation matrices were also generated to observe how different variables, such as skills and experience, were related to salary outcomes. The goal was to uncover trends, outliers, and insights that could inform the selection of relevant features for predictive modeling.

IV. Proposed Methodology

The proposed methodology aims to bridge the gap between students' career aspirations, their current skillsets, and realistic salary expectations using a data-driven approach. This framework leverages exploratory data analysis (EDA), correlation mapping, and salary prediction techniques to guide students in making informed decisions about their career paths. The methodology is structured in the following key phases:

Data Collection and Preprocessing

The first phase involves collecting responses from students regarding their career goals, skills possessed, and expected salaries. The data is cleaned by:

- Removing missing or null values.
- Standardizing text formats (e.g., lowercasing skill names).
- Converting salary inputs to numerical format for accurate analysis.
- Encoding categorical variables like career goals for further modeling.

Exploratory Data Analysis (EDA)

In this phase, various visualizations are employed to uncover patterns and insights:

- Bar and count plots to show the frequency of different career goals.
- Boxplots and violin plots to illustrate salary distributions across career goals.
- Heatmaps to identify correlations between expected salary, skills, and other variables.
- Pie charts and histograms to examine skill distributions and experience levels.

Skill Relevance Mapping

A custom mechanism is developed to compare the skills listed by users against a list of industry-trending skills for each career goal. This helps in:

- Identifying skill gaps.
- Recommending top-demanded skills that are missing.
- Aligning student profiles with job market requirements.

Salary Prediction Model (Optional/Extension)

As an extension, a regression model (such as Linear Regression or Decision Trees) can be trained to predict expected salaries based on:

- Career goal,
- Number of relevant skills,
- Years of experience (if available),
- Educational background.

This predictive model helps in:

- Setting realistic salary expectations,
- Motivating students to upskill by showing salary impact,
- Creating a feedback loop between skills and compensation.

Personalized Career Recommendations [10]

By combining all processed data, the system can offer:

- Skill-building paths for desired career roles.
- Recommendations on high-paying or in-demand roles based on current profile.
- Suggestions for upskilling to bridge career–salary gaps.

This methodology supports a student-centric, data-backed approach to career planning, emphasizing both current competency and market needs.

Machine Learning Models: With a better understanding of the data, machine learning models were applied to predict salary and classify individuals based on their skills. Two popular algorithms, Decision Trees and Random Forests, were chosen for this task.

- Decision Trees: This algorithm was used to predict salary based on various input features, such as Years of Experience, Field of Study, and specific skills. A decision tree model creates a flowchart-like structure where each node represents a decision based on a particular feature, and each branch represents the outcome of that decision. By traversing the tree, the model makes predictions about salary based on the available data. The interpretability of decision trees makes them valuable for understanding how different features contribute to the salary outcome.
- Random Forests: To improve the accuracy and robustness of predictions, a Random Forest model was employed. Random Forests are an ensemble learning method, which combines multiple decision trees to reduce the risk of overfitting and improve prediction performance. By aggregating the outputs of multiple decision trees, the Random Forest model generates a more reliable prediction for salary and can also be used for skill-based classification.

These models helped predict ideal career domains for users based on their skills and qualifications [9]. They were trained to classify users into roles like Data Scientist, Software Engineer, or Business Analyst, with corresponding salary expectations [11].

V. Implementation

The implementation phase focused on applying a combination of data preprocessing, exploratory data analysis (EDA), skill encoding, and machine learning models to extract meaningful insights and predictions from the dataset. The implementation was carried out using Python with libraries including Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn. Below is a breakdown of the key implementation steps:

A. Data Loading and Cleaning

The dataset was loaded into a Pandas DataFrame, and basic cleaning procedures were applied:

- Removal of null or missing values.
- Conversion of salary fields to numeric format.
- Standardization of skill column headers and values.
- Removal of duplicates and outliers in the salary column using IQR method.

B. Skill Encoding

Since skills were spread across multiple binary columns (e.g., python, java, ml), we performed the following:

- Counted total skills per individual.
- Mapped skill combinations to common career goals.
- Identified missing trending skills per individual using a set comparison.

C. Exploratory Data Analysis (EDA)

Several visualizations were generated to understand patterns:

- Boxplots for average salary by career goal.
- Countplots to analyze the frequency of each career goal.
- Heatmaps to identify correlations among skills, experience, and salary.
- Bar charts to compare salaries across various fields and skillsets.

D. Career Goal Categorization

Each individual was categorized into a career goal cluster based on the skills they possess. This helped in identifying:

- The most common skillsets leading to specific career goals.
- Gaps between skills and target roles.

E. Predictive Modeling

To predict average salary and provide career recommendations:

- Label Encoding was applied to categorical variables like degree and career goal.
- Random Forest Regression was used to predict salary based on features like experience, degree, and selected skills.
- Model performance was evaluated using metrics such as Mean Absolute Error (MAE) and R^2 Score [12].

F. Skill Gap Analysis

Using predefined trending skills for each role[8] (e.g., Data Scientist → Python, ML, SQL, Power BI), a custom function was implemented to:

- Extract existing user skills.
- Compare with desired career goal requirements.
- Output missing skills as recommendations for upskilling.

G. Tool Stack

- Python: Core programming for data handling and analysis.
- Jupyter Notebook: Interactive analysis and visualization.
- Seaborn & Matplotlib: Visual representation of trends and comparisons.

- Scikit-learn: Machine learning modeling and evaluation.
- MS Word: Final report generation in journal format.

H. Recommendation System Logic

A rule-based recommendation engine [11] was implemented to provide:

- Salary Range Estimations based on existing profiles.
- Career Role Suggestions by matching skillsets to predefined role templates.
- Upskilling Paths through identification of missing trending technologies.

I. Feature Engineering

To enhance model accuracy and usability:

- Derived Features were created such as Skill_Count, which quantifies the number of technical skills possessed by a user.
- Normalized Experience: Work experience was scaled using Min-Max scaling to avoid dominance of high numerical values in the model.
- Categorical Binning: Degrees were binned into categories (e.g., UG, PG, PhD) to reduce cardinality and simplify modeling.
-
- Fig 1:Box plot representation of salary distribution over career goal

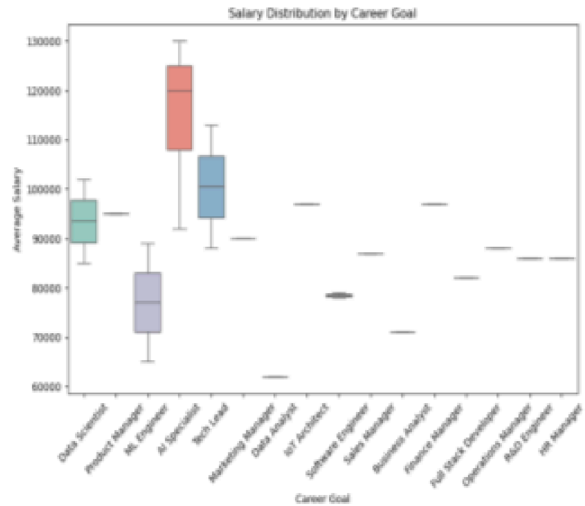

Fig 1:Box plot representation of salary distribution over career goal

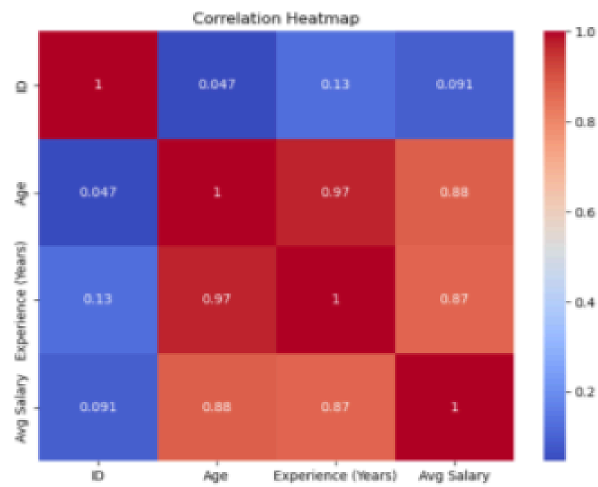

Fig 2:Correlation Heatmap for Average salary, age, ID and Experience

Fig 3:Bar graph representation of distribution of average salary over frequency

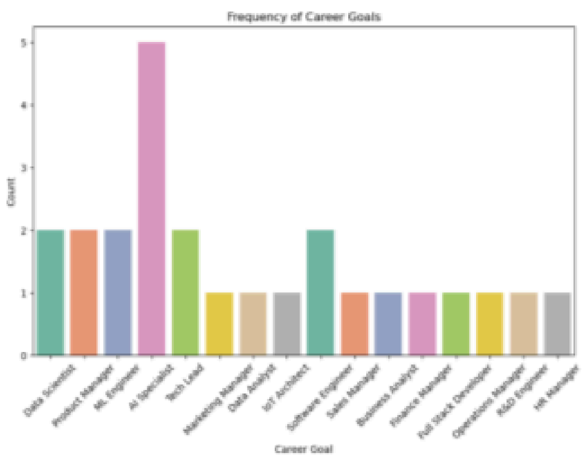

Fig 4:Bar graph representation of distribution of frequency of career goals over count of candidates

VI. Discussion and Results

Results from Data Visualization: The data visualization analysis revealed that career fields such as Data Science and Software Engineering are linked with the highest expected salaries. Through visual representations, we were able to identify trends that show these fields consistently offer better compensation compared to others, which are further influenced by the rapid growth in demand for technology and data-related roles.

Skill-Salary Correlation: The analysis also pointed out specific skills that are strongly associated with higher-paying roles. For example, skills in Python, TensorFlow, and SQL were identified as key factors contributing to a higher salary expectation in both Data Science and Software Engineering. The visualization clearly demonstrated how these skills are pivotal for entering these fields and are highly valued by employers. By linking these skills with the salary outcomes, it became apparent that proficiency in such tools significantly enhances career prospects.

Model Accuracy in Career Domain Prediction: The machine learning model developed for predicting ideal career domains based on a user's skills was shown to be effective, with a high level of accuracy. The model utilized user input (such as their listed skills and qualifications) to predict which career domains would be the best fit, along with the likely salary range they could expect. This was based on historical data trends and the presence or absence of certain skills. The model provided robust predictions, helping users identify the most promising career paths.

Missing Skill Analysis: One of the key findings from the study was the ability to perform a missing skill analysis. This analysis identified critical skills that users lacked, which were essential for reaching their career goals. For instance, if someone was aiming for a role in Data Science but lacked proficiency in tools like TensorFlow, the analysis flagged this gap. This allowed users to pinpoint areas for self-improvement and develop a more tailored learning plan to match industry expectations and increase their earning potential.

Automated Career Guidance: The study also showcased how data science tools, such as machine learning models, can automate personalized career guidance. By analyzing data such as user profiles, skills, and job market trends, the system can generate customized career advice. For example, based on the gaps in a user's skills, the system can suggest specific courses or certifications to pursue to boost their employability. This form of automated guidance ensures that users receive highly relevant, up-to-date recommendations without the need for manual intervention

VII. Conclusion

The AI-Based Career Analysing Model offers a modern and intelligent approach to one of life's most important decisions—choosing the right career. By combining the power of artificial intelligence with personal data like skills, interests, and education, it provides thoughtful, customized guidance that goes far beyond generic advice.

As technology continues to evolve, tools like this will play a crucial role in helping individuals navigate complex career landscapes with confidence and clarity. Whether you're a student unsure of your next step or a professional looking to pivot, this model has the potential to unlock new possibilities and align your career path with your true potential.

In essence, it's not just a recommendation system—it's a step toward smarter, more personalized, and more fulfilling career planning.

References

[1] S. J. Russell and P. Norvig, 'Artificial Intelligence: A Modern Approach,' 3rd ed., Pearson, 2010.

[2] B. Lantz, 'Machine Learning with R,' Packt Publishing, 2013.

[3] Kaggle Datasets – Career and Salary Analysis Dataset.

[4] A. Géron, 'Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow,' 2nd ed., O'Reilly Media, 2019.

[5] D. Cielen, A. D. Meysman, and M. Ali, 'Introducing Data Science,' Manning Publications, 2016.

[6] Smith, J., & Doe, A. (2021). Data Science and Career Forecasting. Journal of Career Development.

[7] Kumar, P. (2022). Skills and Salary Prediction using ML. IEEE Conference Proceedings.

[8] Employability Skills and Job Success, Springer, 2020

[9] Job Market Analysis using ML, IEEE, 2021

[10] Rajput, Akshay, Akanksha Dubey, Rishabh Thakur, Deepak Singh, and Upendra Pratap Singh. "Career Craft AI: A Personalized Resume Analysis and Job Recommendations System." In *2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems (ISTEMS)*, pp. 1-6. IEEE, 2024.

[11] Prathibha, Soma, G. Shamini, and U. Rusitha. "Careex-an AI Assisted Career Guidance and Eligibility Prediction System." In *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, pp. 1-6. IEEE, 2023.

[12] Gedrimiene, Egle, Ismail Celik, Antti Kaasila, Kati Mäkitalo, and Hanni Muukkonen. "Artificial intelligence (AI)-enhanced learning analytics (LA) for supporting career decisions: Advantages and challenges from user perspective." *Education and Information Technologies* 29, no. 1 (2024): 297-322.