



Employment pattern prediction by education level.

Thakur Shyamnarayan Degree College

Guide: Ms. Chhaya Mishra

Ms. Simran Sihan, Ms. Amruta Namdeo More

Undergraduate Students Department of Data Science

University of Mumbai, Mumbai, India

I. Abstract

The increasing competition in modern labor markets has intensified the need to understand the relationship between educational attainment and employment outcomes. Employment patterns are influenced by multiple demographic and academic variables such as education level, academic performance, internship exposure, age, and region of study. Identifying these relationships through statistical analysis provides meaningful insights for students, academic institutions, and policymakers.

This research presents an Exploratory Data Analysis (EDA) study conducted on a structured dataset containing variables including Age, GPA, Education Level, Internship Experience, Region of Study, and Employment Status. The analysis was performed using Python-based statistical libraries including Pandas, NumPy, Matplotlib, and Seaborn. Descriptive statistical measures such as mean, median, variance, standard deviation, quartiles, frequency distribution, and mode were computed to understand variable behavior. Furthermore, graphical visualizations such as count plots and box plots were used to examine relationships between categorical and numerical attributes.

The study reveals observable patterns between higher educational qualifications and improved employment outcomes. Internship experience and academic performance also demonstrate positive associations with employability. The findings confirm that structured statistical analysis can effectively uncover meaningful employment trends and provide a foundation for future predictive modeling.

II. Introduction

The transition from education to employment represents one of the most significant stages in an individual's professional life. In recent years, increasing graduation rates and growing workforce competition have made employability analysis an important area of academic and economic research. While higher education is often associated with better job opportunities, employment outcomes are influenced by multiple interacting factors including academic performance, practical exposure, demographic characteristics, and regional conditions.

Traditional employment studies frequently rely on large-scale labor surveys; however, smaller structured datasets also offer valuable insights when analyzed systematically. Exploratory Data Analysis (EDA) plays a crucial role in understanding data distribution, identifying patterns, detecting anomalies, and interpreting relationships before applying advanced modeling techniques.

This study focuses on analyzing employment patterns using descriptive statistical methods and visualization techniques. By examining how variables such as education level, GPA, internship experience, age, and region of study relate to employment status, the research aims to provide data-driven insights into employability determinants. The approach emphasizes clarity, interpretability, and structured statistical reasoning.

III. Review of Literature

The relationship between education and employment has been widely examined in labor economics and workforce analytics. Human Capital Theory suggests that education enhances an individual's productivity, thereby increasing employability and earning potential. Several empirical studies indicate that higher academic qualifications are positively correlated with employment stability and career advancement.

Research has also emphasized the role of practical exposure, particularly internships, in improving job readiness. Internship experience provides industry exposure, skill development, and professional networking opportunities, which significantly enhance employment prospects. Additionally, academic performance, often measured through GPA, has been found to influence recruitment decisions in entry-level positions.

In recent years, data-driven analytical approaches have been increasingly applied to employment studies. Statistical analysis and visualization techniques allow researchers to uncover patterns and trends within datasets, supporting evidence-based decision-making. However, many studies move directly to predictive modeling without thoroughly examining descriptive relationships. The present study contributes by focusing extensively on structured Exploratory Data Analysis to understand employment behavior prior to model implementation.

IV. Objectives

The primary objectives of this research are:

- To analyze the distribution of employment status within the dataset
- To examine the relationship between education level and employment outcomes
- To study the impact of internship experience on employability
- To analyze GPA variation across employment categories
- To examine age distribution differences between employed and unemployed individuals
- To visualize employment trends using statistical graphs

V. Hypothesis

This study is based on the following hypotheses:

- H1: Higher education levels are positively associated with employment status.
- H2: Individuals with internship experience demonstrate higher employment probability.
- H3: GPA distribution differs significantly between employed and unemployed individuals.
- H0: No significant relationship exists between the selected variables and employment status.

These hypotheses are examined using descriptive statistics and graphical interpretation.

VI. Research Methodology

The research adopts a quantitative analytical approach using Exploratory Data Analysis techniques.

1. Data Collection

The dataset was imported into Python using the Pandas library. The dataset contains structured variables including:

- Age
- GPA
- Education Level
- Internship Experience
- Region of Study
- Employment Status

2. Tools and Technologies

The analysis was conducted using:

- Python Programming Language
- Pandas for data manipulation
- NumPy for numerical operations
- Matplotlib for graphical visualization
- Seaborn for statistical plotting

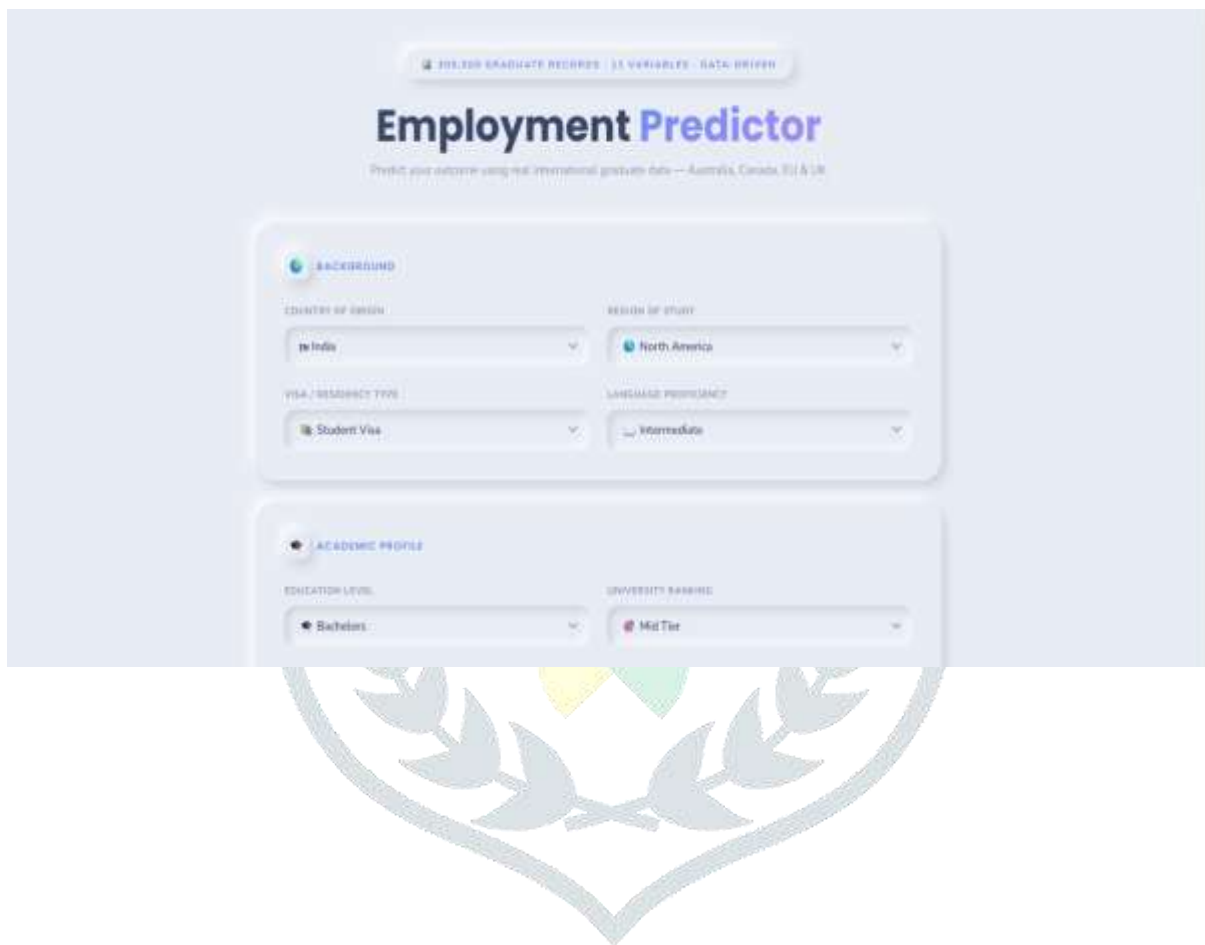
3. Statistical Measures

The following statistical techniques were applied:

- Mean
- Median
- Standard Deviation
- Variance
- Quartiles
- Mode
- Frequency Distribution
- Percentage Distribution

These measures provide insights into central tendency, dispersion, and data spread.

4. UI Visualization



The screenshot displays the user interface of the 'Employment Predictor' application. At the top, a header indicates '305,200 GRADUATE RECORDS | 11 VARIABLES | DATA-DRIVEN'. The main title 'Employment Predictor' is prominently displayed, with a subtitle 'Predict your outcome using real international program data — Australia, Canada, EU & US'. The interface is divided into two main sections: 'BACKGROUND' and 'ACADEMIC PROFILE'. The 'BACKGROUND' section includes dropdown menus for 'COUNTRY OF ORIGIN' (set to India), 'REGION OF STUDY' (set to North America), 'VISA / RESIDENCY TYPE' (set to Student Visa), and 'LANGUAGE PROFICIENCY' (set to Intermediate). The 'ACADEMIC PROFILE' section includes dropdown menus for 'EDUCATION LEVEL' (set to Bachelors) and 'UNIVERSITY RANKING' (set to Mid Tier). A large, stylized logo featuring a shield with a laurel wreath and a central emblem is visible at the bottom of the page.

ACADEMIC PROFILE

EDUCATION LEVEL: Bachelors

UNIVERSITY RANKING: Mid Tier

FIELD OF STUDY: Business

GPA (range 0-4 · step 0.1): 3.2

YEARS SINCE GRADUATION (range 0-8 · dataset split: 0-2yrs=56.8% employed, 3-9yrs=46.0%): 1yr

PERSONAL PROFILE

AGE (range 18-60): 28 yrs

INTERNSHIP EXPERIENCE: None

Predict



VII. Data Interpretation

The descriptive analysis reveals several meaningful patterns.

The GPA analysis indicates measurable variation across individuals. The mean and quartile values suggest moderate dispersion, while boxplot visualization shows that employed individuals tend to cluster around relatively higher GPA values compared to unemployed individuals.

Age distribution analysis highlights that employment appears more concentrated within certain age groups. While age alone does not determine employment, the boxplot suggests variation in median age between employment categories.

Education level distribution demonstrates a noticeable difference in employment frequency across qualification categories. Higher educational attainment corresponds with comparatively higher employment counts, supporting the hypothesis of positive association.

Internship experience analysis indicates that individuals with internship exposure show higher representation in the employed category. This supports the assumption that practical experience enhances employability.

Regional distribution analysis shows variation in employment across different study regions, indicating that geographic factors may influence opportunities.

Overall, the graphical and statistical analysis confirms observable associations among education, internship experience, academic performance, and employment outcomes.

VIII. Significance of the Study

This study provides valuable insights into employment determinants through structured statistical analysis. The findings can assist:

- Students in understanding employability factors
- Educational institutions in designing skill-oriented curricula
- Career counselors in guiding students effectively
- Researchers in developing predictive employment models

By focusing on data interpretation before model implementation, the research emphasizes analytical clarity and informed decision-making.

IX. Scope

The scope of the study includes:

- Descriptive statistical analysis of employment data
- Visualization-based interpretation of relationships
- Identification of key employability indicators

The study provides a foundation for future expansion into predictive modeling and advanced statistical testing.

X. Limitations

Despite meaningful insights, certain limitations exist:

- The study is limited to descriptive analysis
- No predictive machine learning models were implemented
- Dataset size and structure may limit generalizability
- Causal relationships cannot be established

These limitations highlight areas for future research expansion.

XI. Findings

The major findings of the study are:

- Education level demonstrates a positive relationship with employment status
- Internship experience significantly enhances employability
- GPA variation exists between employment categories
- Age and regional factors contribute to employment distribution
- Statistical visualization effectively highlights employment trends

XII. Suggestions

Future improvements may include:

- Implementation of Logistic Regression or Random Forest models
- Inclusion of larger real-world employment datasets
- Application of Chi-square and ANOVA tests
- Development of predictive dashboards
- Integration of skill-based variables

XIII. Conclusion

This study successfully demonstrates the application of Exploratory Data Analysis in understanding employment patterns. By systematically applying descriptive statistics and visualization techniques, meaningful relationships between education level, internship experience, GPA, age, and employment status were identified.

The research confirms that structured data analysis serves as a strong foundation for workforce analytics and future predictive modeling. The findings emphasize the importance of academic qualifications and practical exposure in improving employment outcomes.

The study provides both academic relevance and practical value, offering a comprehensive analytical perspective on employability trends.

XIV. References

- 1. **KaggleDataset:**
Employment Prediction Dataset obtained from Kaggle, used for training and testing the Logistic Regression model.
- 2. **Python(3.x):**
Core programming language used for backend development and machine learning implementation.
- 3. **Scikit-learn:**
Used for implementing Logistic Regression, data splitting (train_test_split), and feature scaling (StandardScaler).

- **4. Pandas&NumPy:**
Used for data preprocessing, cleaning, and numerical computations.
- **5. FlaskFramework:**
Used to build the backend server and connect the machine learning model with the frontend interface.
- **6. HTML5&CSS3:**
Used for designing the responsive user interface and form-based input system.
- **7. GoogleGeminiAPI(GenerativeAI):**
Used for generating intelligent explanations and enhancing prediction interpretation.

