Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Streamlit-Based Application for Predicting Job Prospects After Graduation using Logistic Regression

Hilman Ihza Amrullah [1]*, Taufik Qurhahman[2], Elkin Rilvani[3]

[1,2,3] *Informatics Engineering, Pelita Bangsa University*
hilmanamr22@gmail.com[1*], taufikqurhahman@gmail.com[2], elkin.rilvani@pelitabangsa.ac.id[3]

Abstract

This study presents the development of a web-based application designed to predict the likelihood of university students securing employment after graduation using logistic regression. The application was built with the Streamlit framework and processes input data such as GPA, internships, projects, certifications, soft skills, aptitude test scores, and participation in training or extracurricular activities. The model was trained using a dataset sourced from Kaggle, which reflects various academic and experiential attributes of students. After preprocessing and model training, the logistic regression model achieved an accuracy of 85%, with additional evaluation metrics indicating strong predictive performance. The application features real-time prediction, visual categorization of employment probability, and factor contribution analysis. The results show that aptitude test scores, GPA, and soft skills are the most influential factors in determining employment outcomes. This tool serves both predictive and educational purposes by providing career-related insights for students and decision-making support for academic institutions in planning targeted development programs.

Keywords: *Career Prediction, Employment Readiness, Logistic Regression, Streamlit, Student Profiling*

1. Introduction

In today's digital age, developments in information technology have driven major transformations in various aspects of life, particularly in the fields of education and employment. Information technology is a type of technology used to manage data. This includes activities such as processing, collecting, organizing, storing, and manipulating data in various ways to produce quality information, i.e., information that is relevant, accurate, and timely. This information is used for various purposes, such as personal, business, and government, and serves as strategic information in the decision-making process [1]. The information produced must be relevant, accurate, and timely to support the decision-making process optimally. This quality information plays a crucial role across various sectors, from personal needs, business development, to government policies. In the context of education and employment, information technology has become a bridge between the academic world and the industrial world, enabling the synchronization of competency requirements with the availability of quality human resources.

However, one of the major challenges faced by students today is uncertainty about job prospects after completing their studies at university. Even though they have completed their formal education, many graduates experience difficulties entering the workforce. Various factors contribute to this issue, such as the mismatch between fields of study and industry needs, lack of organizational experience or internships, insufficient soft skills such as communication, leadership, and teamwork, and low proficiency in information technology, which is now a key requirement in almost all job fields. As a result, many graduates are not practically prepared to face the challenges of an increasingly complex and competitive job market. Therefore, a support system is needed that can assess students' potential and readiness for the job market, enabling them to develop self-improvement strategies early on based on relevant indicators.

In response to this issue, we developed a web-based predictive application through a Data Mining practicum. This application was built using the Streamlit framework and designed to help predict students' chances of finding employment after graduation, utilizing logistic regression algorithms as the primary classification method. Logistic regression was chosen for its ability to handle data with two classes or categories, such as classifying between “likely” and “unlikely” to obtain employment. The application works by analyzing various student attributes, such as GPA, organizational experience, involvement in training or certification activities, and technological proficiency. After the data is entered, the model processes it and provides prediction results in the form of probabilities. With this system, students are expected to be more focused in designing self-development steps, while academic institutions can use it as a tool to develop more data-driven career development policies. This approach is expected to bridge the gap between academic achievements and industry needs in a more practical and adaptive manner.

2. Literatur Review

Reference [2] shows an analysis of alumni employment opportunities and the factors that influence them, against the backdrop of rising educated unemployment due to an imbalance between the number of graduates and job opportunities. Using quantitative methods and a survey of 78 alumni from the Mathematics Education Program at Muhammadiyah University of Bengkulu (graduates from 2015–2016), this study applied ordinal logistic regression to test the influence of variables such as gender, age, GPA, place of origin, work experience during university, and interest in the field of work. The results indicate that GPA, work experience during university, and interest in a specific field of work are the dominant factors in determining the speed at which alumni secure employment. These findings serve as an important reference for educational institutions in designing strategies to enhance students' job readiness.

Reference [3] shows the factors causing unemployment among college graduates, with a background of high educated unemployment despite the increasing number of graduates. Using binary logistic regression and 2018 SAKERNAS data, this study analyzes individuals aged 15–64 years with unemployment status as the dependent variable, and independent variables such as age, gender, marital status, household head status, region of residence, and sector of employment. The results show that age, marital status, and household head status reduce the risk of unemployment, while males and those working in the formal sector have higher unemployment rates. These findings highlight the disparity between higher education outcomes and labor market needs, underscoring the need for strategic policies to prepare graduates for the workforce.

Reference [4] shows the role of external employer branding in shaping students' intentions to apply for jobs in the FMCG sector, against the backdrop of fierce competition among companies in recruiting young talent amid Indonesia's demographic bonus. This quantitative study involved 118 final-year students in DIY who are interested in working in the FMCG sector, with logistic regression analysis on the variables of CSR association, CSR participation, and company familiarity as independent variables, and intention to apply as the dependent variable. The results indicate that CSR participation and company familiarity have a significant positive effect on the intention to apply, while CSR association has a negative effect. These findings confirm that companies actively involved in social activities and with a familiar brand image are more attractive to students, making effective external employer branding strategies key to attracting potential young employees.

2.1. Data Mining

Data mining is the process of discovering patterns or interesting information from selected data using specific techniques or approaches. The techniques, approaches, or algorithms used in data mining vary widely. The selection of the appropriate technique or algorithm depends heavily on the objectives and overall process of knowledge discovery in databases (KDD) [5].

2.2. Logistic Regression Algorithm

Logistic regression is a classification technique used to analyze the relationship between a categorical response variable and several continuous predictor variables. The main focus of logistic regression is to estimate the probability or likelihood of an event occurring based on relevant predictor variables. In the steps of logistic regression, information is used to train the model, then the model is evaluated using test data that was not used during the training process [6].

2.3. Streamlit

Streamlit is an open-source framework that uses Python. This framework is designed to help developers create interactive web applications in the fields of data science and machine learning [7].

2.4. Python

Python is a dynamic programming language that is widely used to create applications in various fields. Thus, a program can be written in several ways at once. For example, a graphical interface can be built using an object-oriented approach, while processing can be done in a functional or procedural style. The Python programming language has various features that can be utilized by software developers [8].

3. Research Methodology

3.1. Research approach and methods

This study uses a quantitative approach with applied experimental methods, as it focuses on the development and implementation of a model for predicting employment opportunities using logistic regression algorithms. This model is integrated into an interactive web-based application using the Streamlit framework, which allows users to perform simulations and predictions in real time based on student data.

3.2. Data sources and types

The data used in this study was sourced from Kaggle and is secondary and quantitative in nature. The dataset reflects the condition of students in terms of their readiness to enter the workforce after graduation, with the following attributes:

1. GPA (Grade Point Average, scale of 4.00)
2. Number of Internships
3. Number of Projects
4. Number of Certifications or Workshops
5. Aptitude Test Score
6. Soft Skills Rating
7. Extracurricular Activity Participation (Yes/No)

8. Participation in Placement Training (Yes/No)

The prediction target is PlacementStatus, which is the job placement status (Placed/Not Placed). The input variables include all numerical attributes and the results of categorical variable conversions.

3.3. Data collection and processing techniques

Data is processed through several stages, including:

1. Data Preprocessing
   a. Data cleaning to remove null or invalid values
   b. Conversion of categorical data (such as training and extracurricular participation) to numerical form
   c. Normalizing numerical features such as GPA, Aptitude Test, and Soft Skills using StandardScaler
2. Dataset Splitting
   a. The dataset is split into 80% training data and 20% test data using the train_test_split function from the scikit-learn library
   b. The goal is to train the model on a portion of the data and test it on data that the model has never seen before
3. Model Training
   a. The Logistic Regression model is trained using training data to learn the relationship between input variables and the likelihood of students obtaining employment
   b. Testing is performed on test data to evaluate model performance

3.4. Application design

The application was developed using Python and the Streamlit framework, with the following main features:

1. Interactive input form using sliders and dropdowns
2. Visualization of job opportunity predictions in the form of percentages and categories (High, Moderate, Low)
3. Table of main factor contributions, showing the influence of each attribute on the prediction results
4. imple and intuitive display, so that it can be easily used by students and campus official

3.5. System flow

Here is the workflow of the application system:

1. Users upload or input student data via a form
2. The system performs data preprocessing as done during model training
3. The Logistic Regression model predicts the probability of getting a job based on the input
4. The results are displayed as a percentage probability, accompanied by prediction categories such as “Very High Probability,” “Moderate,” or “Low.”
5. The application also displays the contribution of each factor and self-development suggestions based on the prediction results.

4. Results and Discussion

4.1. Application interface

This application provides an interactive interface for predicting the chances of getting a job after graduating from college, with the following main features:

1. Average data: The system displays the average contribution of each factor (e.g., GPA, internships, projects, certifications, soft skills, etc.) to employment opportunities, based on a logistic regression model trained on a dataset of students from Kaggle.
2. Manual input: Users can manually adjust the value of each factor using sliders and dropdown menus, e.g., GPA (on a 4.00 scale), Number of internships (0–10), Number of projects (0–10), Certifications/Workshops (0–10), Aptitude Test Score, Soft Skills Rating (0.0–5.0), Extracurricular Activity Participation (Yes/No), and Placement Training (Yes/No).
3. Prediction: After the inputs are set, the application will calculate and display the probability of employment opportunities (e.g., 78.5%) using a logistic regression model.
4. Categories: The system automatically groups prediction results into several categories based on the percentage of employment opportunities, namely High (80% and above), Fair (60% to 79%), Moderate (40% to 59%), and Low (below 40%). Each category is accompanied by a different color notification for easy identification, as well as relevant self-development suggestions to help users improve their employment prospects.
5. Visualization: Prediction results are visualized with a progress bar, employment probability percentage, and factor contribution table, making them easy for users to understand.

Faktor-Faktor Penentu Peluang Mendapatkan Kerja

Faktor	Rata-rata Data	Koefisien Model	Kontribusi
Tes Aptitude	79.45	0.08	6.65
IPK (skala 4)	3.08	1.15	3.53
Soft Skills	4.32	0.78	3.38
Proyek	2.03	0.36	0.73
Pelatihan Penempatan	0.73	0.96	0.7
Ekstrakurikuler	0.59	0.97	0.57
Sertifikasi/Workshop	1.01	0.2	0.21
Magang	1.05	0.04	0.04

Fig. 1: Predicted Employment Outlook and Factors Based on Average Data

Fig. 1 shows the initial display of the application for predicting the likelihood of getting a job after graduating from college. At the top, it is explained that the model used is Logistic Regression, which analyzes the influence of various factors such as GPA, number of internships, projects, certifications, soft skills, and placement training on the likelihood of students getting a job. Each feature has a weight or coefficient that represents its contribution to the prediction result; the larger the coefficient value, the more significant the feature's influence in increasing employment opportunities.

Below the model explanation, the application presents a table displaying a complete list of variables with their average values, model coefficients, and the contribution of each feature. This contribution is calculated by multiplying the average data value by the regression coefficient. From this visualization, users can immediately identify the most dominant factors, such as Aptitude Tests, GPA, and Soft Skills, which have the greatest contribution to the prediction. With this information, students can focus more on developing the aspects that most influence their job readiness.

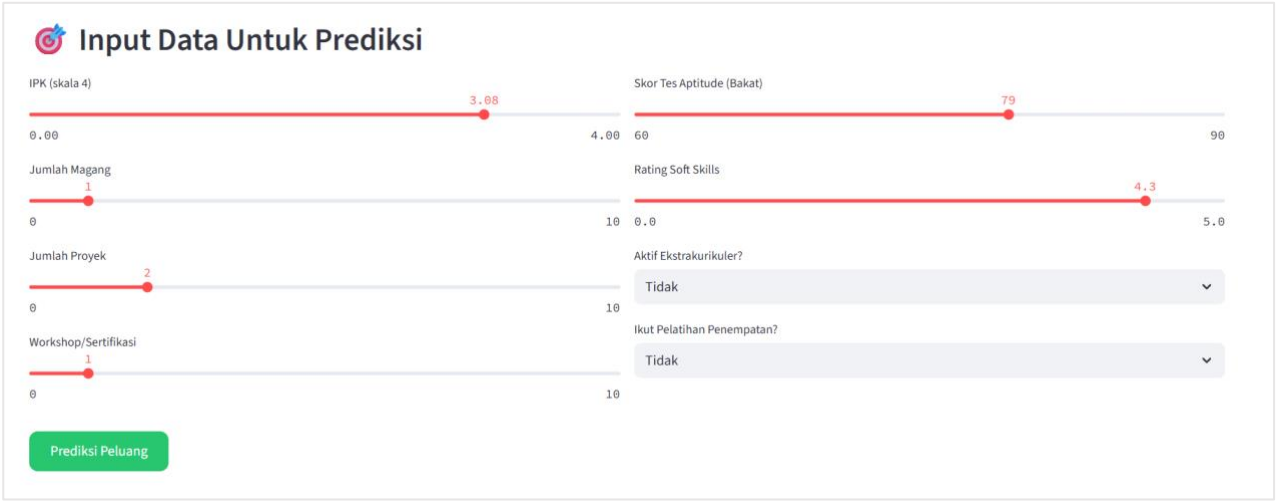

Fig. 2: Manual data input display

Fig. 2 shows the manual input feature in the application for predicting job opportunities after graduation. Users can enter values for key factors such as GPA, number of internships, projects, certifications, aptitude test scores, soft skills, extracurricular activities, and placement training using interactive sliders and select boxes. This interface is designed to be user-friendly, allowing users to adjust values based on actual conditions or run simulations of various scenarios to see their impact on prediction results.

With a responsive two-column design, the display remains neat and efficient even when loading many variables. After all data is entered, users simply click the Predict Opportunities button to view real-time prediction results. This feature not only helps users understand the direct impact of each factor but also serves as a concrete tool for planning personal development to enhance employment prospects after graduation.

Fig. 3: Prediction Results

Fig. 3 shows the final results of the job opportunity prediction after the user has filled in all the input data on the application form. The system automatically calculates the probability using a Logistic Regression model and displays the results in a clear visual format, such as a progress bar and job opportunity percentage. Additionally, the results are categorized into categories such as very high, moderate, or low, complete with color notifications and icons to help users understand the level of probability they have.

More than just displaying numbers, the application also provides relevant **narrative recommendations** based on the prediction results. If the probability is high, users will receive motivational messages, while if it is moderate or low, the system provides self-development suggestions such as improving soft skills or adding internship experience. This display is not only interactive and informative, but also serves as an educational tool and practical guide to help students understand and improve their readiness to enter the workforce.

4.2. Model evaluation

The Logistic Regression Model was evaluated using the following key metrics:

a. Accuracy: 85%
b. Precision: 83%
c. Recall: 80%
d. F1-score: 81%
e. ROC-AUC: 0.88

Interpretation:

a. An accuracy of 85% indicates that the model is capable of predicting students' job placement status with a high degree of accuracy.
b. High precision means that most of the “Placed” predictions are indeed students who have successfully secured employment.
c. A recall of 80% indicates that the model is sufficiently sensitive in detecting students who have actually been placed in jobs.
d. A balanced F1-score indicates that the model performs stably between accuracy and sensitivity.
e. ROC-AUC approaching 1 indicates that the model is very good at distinguishing between students who are placed in jobs and those who are not.

4.3. Prediction visualization

The prediction results on this app are displayed concisely and clearly through a progress bar that shows the percentage chance of getting a job. Large probability figures, such as 98.8%, are immediately visible below, accompanied by a corresponding notification for example, Very high chance of getting a job! if the prediction is above 80%. This visualization makes it easy for users to instantly understand how high their job placement chances are, complete with motivational messages and a note that the results are based on data-driven simulations.

4.4. Model coefficient

Here are some coefficients from the regression model that show the magnitude of the feature's influence on stress:

Table 1: Regression Coefficients and the Influence of Features on Stress

Factors	Model Coefficient	Contribution
Aptitude Test	0.008	6.65
GPA (Scale 4)	3.08	3.53
Soft Skills	4.32	3.38
Project	2.03	0.73
Placement Training	0.73	0.7
Extracurricular	0.59	0.57
Certification/Workshop	1.01	0.21
Internship	1.05	0.04

The following is an interpretation of Table 1:

a. A positive coefficient for a factor (e.g., GPA, soft skills, projects, etc.) indicates that the higher the value of that factor, the greater the student's chances of finding employment after graduation. In other words, this factor contributes directly to increasing employment opportunities.
b. A higher contribution value indicates that the factor has a stronger influence on average across all student data. Contribution is calculated by multiplying the average factor value by the model coefficient, so the factor with the highest contribution value becomes the primary determinant in predicting employment opportunities.

c. In the table, Aptitude Test has the largest contribution (6.65), followed by GPA (3.53) and Soft Skills (3.38). This means that the average scores of students' Aptitude Test, GPA, and Soft Skills are highly determinative of the prediction results—the higher the scores, the higher the employment prospects.
d. Other factors such as Projects, Placement Training, Extracurricular Activities, Certifications/Workshops, and Internships also have positive coefficients, but their contributions are smaller. This means that these factors still enhance employment prospects, but their influence is not as significant as the three main factors mentioned above.

All coefficients in this table are positive, so there are no factors that reduce employment prospects in this model. If there were negative coefficients, it would mean that an increase in the value of that factor would actually reduce employment prospects (but there are no negative coefficients in your model).

Table 1 shows the results of logistic regression analysis in the form of coefficients for each factor that influences the probability of students obtaining employment after graduation. These coefficients indicate the direction and magnitude of the influence of each independent variable on the predicted probability of employment. A positive coefficient value indicates that the higher the value of that factor, the greater the likelihood of students being hired. For example, GPA has a coefficient of 3.08 and Soft Skills a coefficient of 4.32, indicating a fairly strong influence on the prediction results. Although Aptitude Test has a small coefficient (0.008), this factor remains the largest contributor because its average value is high in the dataset.

Other factors such as the number of projects, internship training, extracurricular activity, certifications, and internships also have positive coefficients, indicating that improving these factors also increases employment opportunities, although their influence is not as significant as GPA or soft skills. Contribution values are calculated by multiplying the average feature value by the model coefficient. This interpretation helps identify the most influential variables, enabling students to focus more on developing key aspects such as GPA, soft skills, and Aptitude Test scores to enhance their job readiness after graduation.

5. Conclusion

Based on the results of the design and implementation of the application for predicting the likelihood of obtaining employment after graduation using the Logistic Regression algorithm, it can be concluded that this application successfully performs all of its functions well. The application is capable of processing student datasets, training predictive models, and displaying evaluation results and predictions interactively through an easy-to-use interface. The prediction process is effective and accurate, as indicated by the high model evaluation scores, which suggest that logistic regression can accurately explain and predict students' employment prospects.

Additionally, the use of the Streamlit framework plays a significant role in simplifying user interaction. Through input features such as sliders and select boxes, students and campus officials can easily perform prediction simulations based on the data entered. Overall, this application is not only useful as a prediction tool but also has the potential to serve as a supportive tool in students' career decision-making. This application can also assist educational institutions in designing more targeted and data-driven self-development programs to enhance students' readiness for the workforce.

References

[1] "Asmawi, Syafei, and Muhammad Yamin, 'Pendidikan Berbasis Teknologi Informasi dan Komunikasi,' Prosiding Seminar Nasional Pendidikan Program Pascasarjana Universitas PGRI Palembang, vol. –, pp. 50–55, May 2019."
[2] M. Mahyudi, "PELUANG ALUMNI PENDIDIKAN MATEMATIKA FKIP UMB DALAM MENDAPATKAN PEKERJAAN DENGAN MENGGUNAKAN ANALISIS REGRESI LOGISTIK," *MEDIA Stat.*, vol. 10, no. 2, hlm. 85, Des 2017, doi: 10.14710/medstat.10.2.85-94.
[3] V. Astriani dan R. Nooraeni, "DETERMINAN PENGANGGURAN LULUSAN PERGURUAN TINGGI DI INDONESIA TAHUN 2018," *J. Pendidik. Ekon. JUPE*, vol. 8, no. 1, hlm. 31–37, Apr 2020, doi: 10.26740/jupe.v8n1.p31-37.
[4] D. Oktarika dan G. Rizky, "Daya Tarik Eksternal Perusahaan Terhadap Niat Melamar Bagi Mahasiswa Tingkat Akhir," *Value J. Manaj. Dan Akunt.*, vol. 19, no. 1, hlm. 385–398, Apr 2024, doi: 10.32534/jv.v19i1.5393.
[5] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, hlm. 213–219, Feb 2017, doi: 10.22202/ei.2016.v2i2.1465.
[6] N. Fitriyani, D. R. Amalia, H. H. Handayani, dan A. F. N. Masruriyah, "Aplikasi Berbasis Web Berdasarkan Model Klasifikasi Algoritma SVM dan Logistic Regression Terhadap Data Diabetes," vol. 7, 2023.
[7] A. Putranto, N. L. Azizah, dan I. R. I. Astutik, "Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit," vol. 4, no. 2, 2023.
[8] D. A. Budi, "Perancangan Sistem Login pada Aplikasi Berbasis GUI Menggunakan Qtdesigner Python," *J. SIMADA Sist. Inf. Dan Manaj. Basis Data*, vol. 4, no. 2, hlm. 92–100, Nov 2021, doi: 10.30873/simada.v4i2.2961.