

## INTISARI

### **APLIKASI MODEL PREDIKSI GAJI PROFESI DATA BERBASIS WEBSITE MENGGUNAKAN FLASK DAN METODE *DECISION TREE***

Wanda Listathea Putri

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Islam Indonesia

Zaman sekarang, informasi mengenai lowongan pekerjaan sudah banyak tersedia secara *online* melalui berbagai situs, di antaranya adalah LinkedIn dan JobStreet. Kedua situs ini termasuk dalam situs lowongan kerja terpercaya dan paling banyak diakses oleh calon pekerja Indonesia. Dalam situs tersebut, tersedia informasi dan kualifikasi apa saja yang dibutuhkan dalam suatu profesi, termasuk informasi gaji yang ditawarkan. Kondisi ideal dalam menentukan gaji seorang pekerja adalah disesuaikan dengan kompetensi dan kualifikasinya. Namun tak jarang, perusahaan menyamaratakan gaji suatu profesi padahal kualifikasinya berbeda. Di sisi lain, calon pekerja juga terkadang kebingungan dan bahkan menjawab tidak disesuaikan dengan kualifikasi yang dimiliki jika ditanyakan mengenai gaji yang diharapkan pada proses negosiasi gaji. Penelitian ini bertujuan untuk memprediksi gaji profesi data yaitu *Data Engineer*, *Data Analyst*, dan *Data Scientist* dengan menggunakan metode terbaik yang diperoleh dari 3 metode yaitu *Random Forest*, *Support Vector Regression*, dan *Decision Tree*, serta menyajikannya dalam bentuk *website* sehingga mudah digunakan. Pengambilan data dilakukan dengan cara *web scraping* pada tanggal 4 November 2021 untuk data LinkedIn dan tanggal 6 Desember 2021 untuk data Jobstreet. Variabel-variabel yang berhubungan cukup kuat terhadap “gaji” adalah “nama profesi data”, “lokasi perusahaan”, “tingkat pekerjaan”, “lama pengalaman”, “ukuran perusahaan”, dan “industri”. Metode terbaik yang diperoleh dalam penelitian ini adalah *Decision Tree* menggunakan SMOTE dengan jumlah data per kategori “nama profesi data” adalah  $n=87$ . Parameter yang digunakan pada metode tersebut merupakan parameter terbaik yang diperoleh dari proses *hyperparameter tuning* dengan hasil evaluasi adalah  $R^2$  sebesar 0,669, RMSE sebesar 147,817, serta MAPE sebesar 14,1%. *Website* hasil *deployment* dari model tersebut dapat diakses pada URL <https://prediksi-gaji-profesi-data.herokuapp.com/>.

**Kata Kunci:** *Deployment*, *Prediksi Gaji*, *Decision Tree*, *SMOTE*, *Web Scraping*

## ABSTRACT

### APPLICATION OF DATA PROFESSION SALARY PREDICTION MODEL BASED ON WEBSITE USING FLASK AND DECISION TREE METHOD

Wanda Listathea Putri

Department of Statistics, Faculty of Matematics and Natural Sciences  
Universitas Islam Indonesia

*Nowadays, information about job vacancies is widely available online through various sites, including LinkedIn and JobStreet. These two sites are included in the most trusted job vacancy sites and are most accessed by prospective Indonesian workers. The site provides information and what qualifications are needed in a profession, including information on the salary offered. The ideal condition in determining a worker's salary is that it is adjusted to his competence and qualifications. But not infrequently, companies generalize the salary of a profession even though the qualifications are different. On the other hand, prospective workers are also sometimes confused and even answer that they are not adjusted to their qualifications when asked about the expected salary in the salary negotiation process. This study aims to predict the salary of the data profession, namely Data Engineer, Data Analyst, and Data Scientist using the best method obtained from 3 methods, namely Random Forest, Support Vector Regression, and Decision Tree, and presents it in the form of a website so that it is easy to use. Data retrieval is done by web scraping on November 4, 2021 for LinkedIn data and December 6, 2021 for Jobstreet data. The variables that are strongly related to "salary" are "name of profession data", "company location", "level of employment", "long experience", "company size", and "industry". The best method obtained in this study is Decision Tree using SMOTE with the amount of data per category "name of data profession" is  $n=87$ . The parameters used in this method are the best parameters obtained from the hyperparameter tuning with the evaluation results being  $R^2$  of 0.669, RMSE of 147.817, and MAPE of 14.1%. The website that is the result of the deployment of the model can be accessed at the URL <https://prediksi-gaji-profesi-data.herokuapp.com/>.*

**Keywords:** *Deployment, Decision Tree, SMOTE, Salary Prediction, Web Scraping*