Software Developer Salary Predictor using Stake-overflow survey 2022

About Survey:

In May 2022 over 70,000 developers told Stack-overflow how they learn and level up, which tools they're using, and what they want. You can get data-set from kaggle or stake-overflow.

Survey focused on below topics:

Developer Profile / Respondent's profile:

- Education.
- Learning to code.
- Experience.
- Developer roles.
- Key territories.
- Demographics.

Technology: Tools and technologies developers are currently using and what they want to use.

Work: Respondent's employment, company info, salary, influence, hobby etc.

Community: How people use Stack Overflow and how connected they feel to the Stack-overflow community.

Professional Developers: Respondent's Work lifestyle, work experience etc.

Methodology: Feedback on survey.

Problem Statement:

Phase I: Analysis

- In first phase, perform analysis to get basic understanding and insights of the data-set. Analysis includes the answers of the following questions
- Analysis and Visualize the Educational background of all the respondents(Professional Developers & Learning to code)
- Learning to code is different for every Respondents like online courses, college degrees etc. Analysis learning to code of all respondents by resource and age.
- Describe top ten online resources used by respondents to "Learn to code".
- Analysis the best online platforms for "Learn to code".
- Describe the average coding experience of all the respondents for top ten countries.

- Analysis and visualize the developer types of all the respondents.
- Describe top ten countries by respondents.
- Analysis and describe the percentage value of age for all the respondents.
- Analysis and describe the top technologies, databases tools, could platforms, web frameworks, libraries, deployment tools, development environments and operating systems used by all the respondents.
- Analysis and visualize the top technologies, databases tools, could platforms, web frameworks, libraries, deployment tools, development environments and operating systems wanted by all the respondents after five years.
- Analysis and visualize the top paying technologies, databases tools, could platforms, web frameworks, libraries, deployment tools and development environments used by all the respondents.
- Analysis and visualize the change in salaries between 2021 and 2022 of top paying technologies, databases tools, could platforms, web frameworks, libraries, deployment tools and development environments used by all the respondents.
- Display the employment status of each respondent and also by geography, and working environment.
- Analysis and Visualize the salary by developer type as well as by geography and by experience.

Phase II: Model Building

In this phase, build a model and evaluate their performance to predict the salary of a person on the basis of his/her education, experience and country.

Phase III: Model Development

Build a web App using flask library to display the prediction of the salary for a person and describe the analysis of a country using top technologies by salary, top developers types by salary etc

Work-flow for Industry Projects:

- System Architecture
- High Level Design
- Component Selection
- Low Level Design
- Core utility design
- Deployment Architecture
- Multistage pipeline for CI/CD
- ML Pipeline Understanding
- Training Pipeline Implementation
- Inference Pipeline Implementation
- Retraining Pipeline Implementation
- Deployment of ML Pipeline on Cloud
- Monitoring of System and Model Performance

Work-flow for this Project:

Problem Statement/Definition:

Types of Data:

Structure and unstructured

Goal:

Decision Making, Prediction or Pattern Discovery.

Machine Learning Types:

Supervised, Unsupervised or Reinforcement Learning.

Type of Problem:

Classification or Regression

Data Gathering:

Kaggle

Data Preparation/Data Cleansing/Feature Engineering:

Imbalanced Data, Feature Extraction and selection, Dimensionality reduction

Data Analysis/EDA:

Stats, Transform, Visualize

Model Training/Model Building:

Feature selection , Model Selection

Model Evaluation and Tuning:

Under-fitting, over-fitting, Trade-off, High Bias, High Variance Confusion Matrix, ROC etc.

Model Deployment:

Flask, Docker, Heroku, Github actions etc.

Model Experiment and Monitoring:

flask_monitoringdashboard

Project Approach:

- ➢ Understanding the dataset
- ➢ Preparing Dataset And Basic Analysis
- ➢ Preparing Dataset For Model Training
- ➢ Training The Model
- ➢ Performance Metrics
- ➢ Prediction Of New Data
- ➢ Pickling the model file
- ➢ Setting Up Github And VS Code
- ➢ Tools And Software Required
- ➢ Creating A New Environment
- ➢ Setting up Git
- ➢ Creating A FLASK Web Application
- ➢ Running An Testing our application
- ➢ Prediction From Front End Application
- ➢ Procfile for Heroku Deployment
- ➢ Deploying The App To Heroku
- ➢ Deploying The App Using Dockers