

Ради Стефчев Радев

8MI3400507

Цел на приложението

- Приложението представлява уеб базирана система от услуги за анализ и прогнозиране на възнагражденията на софтуерните инженери на база на различни характеристики
- Състои се от две част:
 - Подготовка на данните
 - Streamlit web app:
 - Проучване на данните
 - Прогнозиране

Анализ на изискванията към приложението

- Функционални изисквания:
 - Системата сваля данните от интернет(wget, http), разархивира ги и трие ненужните файлове
 - От данните се извличат само необходимите за анализа характеристики, почистват се, етикетират се(labeling) и на база на това се разработва модел, синтезираните данни се записват във формат подходящ за използване от приложението(*.pkl)
 - Уеб приложението се състои от две части, едната визуализира анализ на различните характеристики, а другата прави предположения за очакваното възнаграждение на база на тях

- Нефункционални изисквания:

- Голяма част от логиката на приложението се визуализира в уеб среда, като може да бъде стартирано както в контейнер(***sh build.sh && sh run.sh***) така и директно от командния ред(***pip install -r requirements.txt && streamlit app.py***)
- За стартиране е необходима среда с браузър и работещ docker, използвани са технологии с отворен код.
- Необходимо е данните да бъдат налични (***sh /data/get_data.sh***)

Обосновка на използваните технологии

- Програмни технологии:
 - Python3.11 - предвид широката му употреба за работа с данни и машинно самообучение и големия набор от библиотеки е подходящ избор за реализиране на проекта
 - Bash/Shell Scripting - позволява лесно и бързо, чрез единични команди да се свалят данните, да се разархивират, както и да се извършват различни манипулации върху тях (местене, триене, преименуване)
 - Streamlit - библиотека, позволяваща бързо и лесно да се визуализират данни и модели
 - Pandas - популярна библиотека за анализ и манипулация на данни
 - Scikit-learn - трениране на модела, лек и удобен предвид данните и целите
 - Docker - контейнеризация на системата, позволяваща стартирането и в различни среди и операционни системи

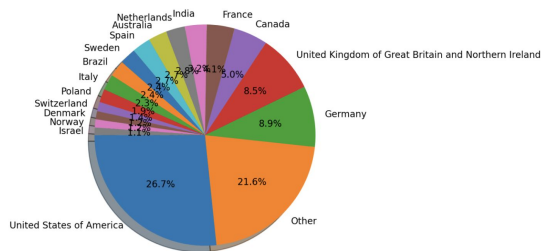
Проектиране на услугите и приложението

- Приложението се състои от 3 модула
 - explore.py - анализ на данните и визуализация на характеристиките

Explore Software Engineer`s Salary

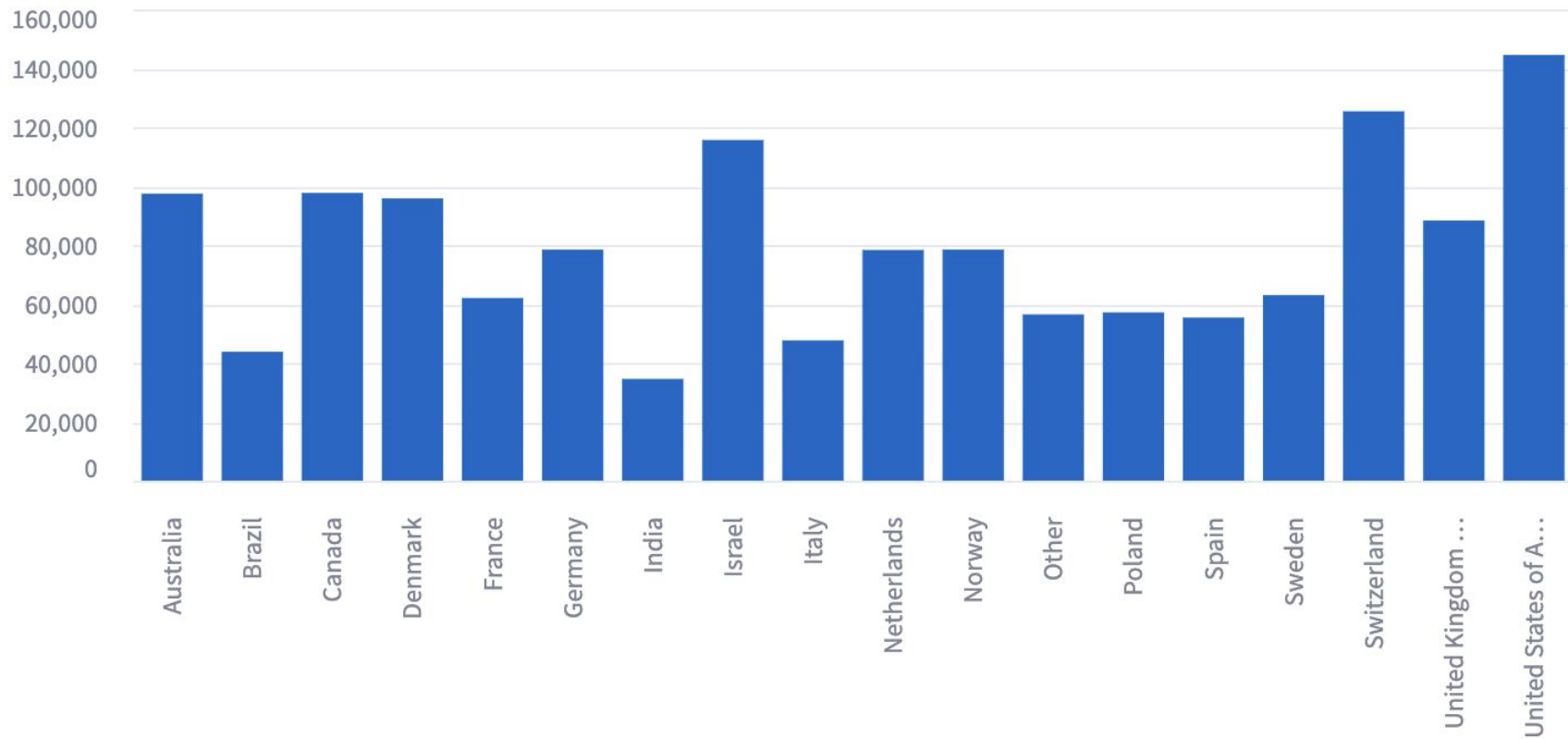
Based on Stack Overflow Developer Survey 2023

Countries

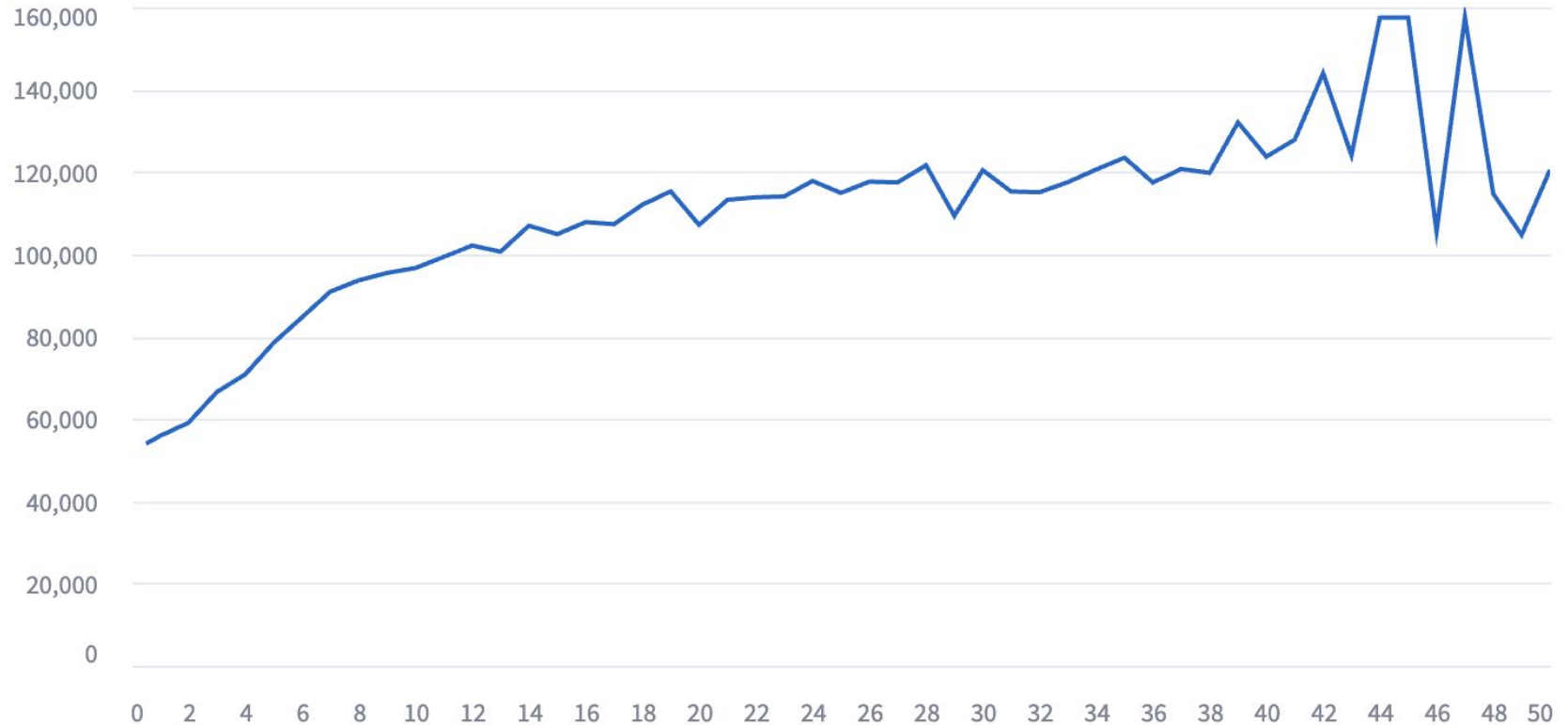


Mean Salary per country

Mean Salary per country



Mean Salary based on experience(years)



Проектиране на услугите и приложението

- prediction.py - изчисляване на прогнозно възнаграждение на база на събраните данни
- apply.py - стартира приложението, навигиране между услугите

Software Developer Salary Prediction

information for the prediction

Country:

United States of America



Degree

Bachelor's degree



Experience (Years):

3

0

50

Predict Salary

The estimated salary is 118482.572 \$ per year

Проектиране на услугите и приложението

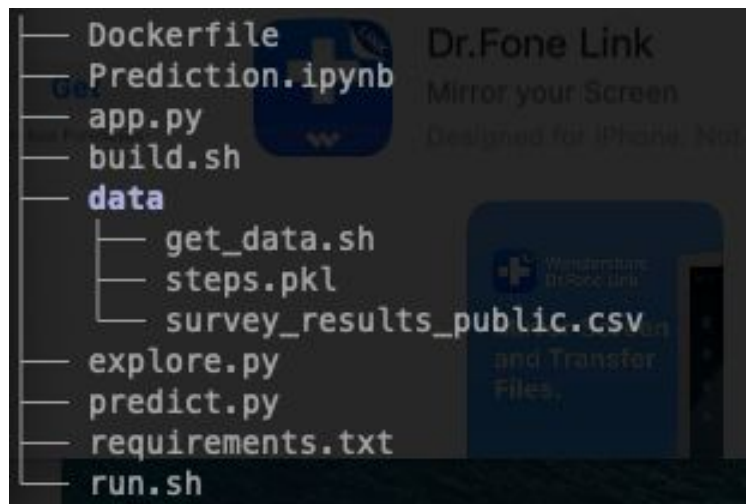
- Други

- Шел скриптове за run, build на проекта и сваляне на суровите данни
- Dockerfile
- Prediction.ipynb - почистване на данните, анализ и трениране на моделите
- requirements.txt

- Протоколи

- Основният протокол за реализацията на приложението е HTTP

Описание на програмната реализация



Описание на програмната реализация

- Системата се състои от три модула, реализирани като две отделни услуги и модул, позволяващ визуализация в браузъра
- Реализирани са като се вика обект от `streamlit`, библиотеката за визуализация, а данните се манипулират с функции(почистване, вадене на отклонения и др.)
- Използваните данни се съхраняват в `csv` и `pkl` формат