AJIT WANKHEDE

(+91) 9284142344

e-mail: ajitwankhede77@gmail.com

Add: Pune, India

LinkedIn | GitHub

WORK EXPERIENCE

Data Science Instructor (Present) Oct 23- Present

Bizpathshala

As an instructor of DS, my responsibilities include curriculum development, teaching, evaluation, mentoring, and research. I also collaborate with faculty members and industry professionals, communicate with non-teaching teams and stakeholders.

Role & Responsibility:

- Teaching Statistics and probability, Mathematics, **Data wrangling** and DBMS, **Machine Learning, Deep learning**, **Artificial Intelligence**, **NLP**, **Computer Vision**, **public speaking.**
- Programming: **Python**, **SQL**

Data Science Research (Present) Apr 23- Present

Agriculture and Sustainable Agro-Economy

The research will address agricultural and Agri-economy issues using state-of-the-art ML, AI models and econometric techniques.

- Develop app using **LLM (GPT 3.5)** and **LangChain** for augmented data annotation task
- Apply **FASSI Similarity Search** method, use **Flat index** method
- **EDA**, **feature engineering** using Python libraries like **NumPy**, **Pandas**, **Seaborn**, **Matplotlib**
- **Statistical methods**, Applied **Monte Carlo** simulation procedure for farm store, Store has experienced a **revenue growth** of **26%** Demo, Defining data privacy policies and preparing research reports

Ai INTERN (6 Months) Sept 22– Feb 23

Venturit consulting services Pvt. Ltd. Pune

Avidhrt is a medical device that measures personal ECG, Pulse Oximetry and Body Temperature. Avidhrt Ai model had 72% Positive Prediction Accuracy on testing data, need to improve above 92% for released.

- Actively involved in daily standup calls and task assigned on Slack and Jira
- Analyzed noisy data and performed **PCA**, **filtration** on 10k+ rows of **clinical data**
- Performed **Feature Engineering** on Bio clinical data using python libraries like **NumPy**, **Pandas**, **Seaborn**, **Matplotlib**, **SciPy**
- To increase the accuracy of the model, applied **Reverse Engineering**, **Data analysis** technic, Fix **Data leakage**, **fine tune XGBoost Classification model**, Train Test and Validate model
- Contribute to maintaining production **CI/CD life cycle** on live servers
- Implemented new parameter that improves model performance from 72% accuracy to **98%** and on real-life clinical datasets got accuracy of **100%**
- Documented research for FDA approval of product

MACHINE LEARNING INTERN (6 Months) Feb 22 – July 22

Rikaian Technology Pvt. Ltd, Pune

Rian is a multinational technology company providing multilingual localization solutions, they wanted to create their own English Japanese language tokenizer and translation model

- Actively involved in daily standup calls on Microsoft team
- Developed **Eng-Jap language Lattice base tokenizer** using **Viterbi Algorithm**
- Fine tune pretrain **Transformer model** on English Japanese Translation database using **Transfer Learning methods**
- Develop application for **NLP Word alignment** problem, Optimize costing of model training

PROFESSIONAL SKILL

- **Tech skill** – Python, **Vector DB**, SQL, **Transformer**, **GPT**, **Xgboost**, ML and DL, A/B testing, Docker, **Facebook AI Similarity Search (FASSI), Chroma DB**
- **Python package** – Numpy, pandas, **Langchain**, **TensorFlow**, **Scikit-learn**, **SciPy**, NLTK, PyTorch, Flask
- **Cloud services** – AWS, Remote Server
- **Work Env Soft.** – Anaconda, VSCode, Jupyter notebook, JIRA, Slack, Microsoft Teams
- **Domain** – Machine learning, Deep learning, Mathematics, Statistics, Data Structure, Clinical and agriculture

EDUCATION

Master of Data Science (CGPA- 9.40) Jan 2021 - July 2022
Fergusson College, Pune

Bachelor of Science, Physics major (73.78%) Mar 2017 - Jun 2020
Sir Parashurambhau College, Pune

PROJECT

Research Paper Summary app:

When it comes to studying a research paper document, it is a very tedious task, and if paper does not meet our needs, we end up with waste of time, so here I have **developed LLM powered Que-Ans app**, answers your question with Ai touch

- **Finetune GPT 3.5 model** on user input data
- **Langchain** framework use to develop application
- **Word embedding**, **Transformers model**, **OPENAi embedding** and model are use
- **Vector database** management tool are use like **Facebook AI Similarity Search (FASSI)**
- Tech skill: Python, Streamlit API, OPENAi API, Prompt engineering
- Packages: **Langchain**, **stremlit**, **pickel**, **pypdf2**, **LLM**, **openai**, **openAIembeddings**

Sentiment Analysis Using Transformers:

The purpose of this activity is to detect hate speech in tweets.

- Packages: **Pandas**, **NumPy, TensorFlow**, **Tranformers**, **Torch (PyTorch), Matplotlib**
- Dataset: TSATC Twitter Sentiment Analysis Training Corpus
- Model: **DistilBert** for **Sequence Classification (BERT Model)**
- Model specification: Vocab size = 30522, **Positiog Embeddings**= 512, Hidd. Dim = 3072
- Activation function: **'Gelu'**
- Tokenization, padding and encoding of data: DistilBertTokenizerFast
- Accuracy: 92%