Pratik Patil

Data Scientist

Total of 3 years relevant exp in data science ML/NLP field and handling large datasets with knowledge in healthcare, Trading sectors, Banks customer, student management, Insurance domain.

priyanshuagri@gmail.com | 8778687149 | Thane, Maharashtra, India

WORK EXPERIENCE

Data Scientist

ACL Digital

08/2021 - Present,

Achievements/Tasks

- Implemented **NLP (bert, roberta etc.), SVM model** which predicts multiple class for customer details, document class, information extraction, predictive modelling**.**
- Handled **structured and Unstructured** English/Chinese/Japanese customers, bank and business NLP data for PayPal client and performed **data labeling using classification, web scraping, cleaning** pipeline on docker container.
- Researched on **optimizations** for training large models on **Google cloud platform**, with less memory consumption/time usage and improved accuracy for models.
- Performed python **Data visualization/analysis, NLP semantic and sentiment** analysis created Regex, and **linear regression model for PayPal** transaction money data.
- Worked on creating custom conversational domain data using **Machine translation**, **Data Augmentation**, **ngram for Chinese/Japanese lang.**

NLP scientist

Genentech

02/2021 - 07/2021,

Achievements/Tasks

- Developed and deployed **clustering classification, K-NN(**K-Nearest Neighbors**) bert model** for NER, documents tagging from trading sectors, medical sector, resume.
- Handled **data creation, cleaning, and analysis/visualization** for large datasets.
- Created deep learning model for **Checkbox detection** in documents which include **RCNN Resnet50 classification model**.
- Worked on Python tools **Part of Speech tagging, Semantic analysis** for trading entities.

NLP engineer

Taiger

10/2019 - 11/2020,

Achievements/Tasks

- Developed **extraction methods, utilized SVM classification for documents, topic modelling**.
- Accomplished **data collection, cleaning and data analysis, data/text mining** for large dataset from Gov, medical sector, Insurance company, Visa Centre.
- Trained deep learning model for **Cursive Handwritten Text Recognition** in documents with **Faster RCNN Resnet50** learned fine tuning model, handling hyperparameters.
- Developed python **Selenium web scraping** of Global Flights data etc. Worked on
- **Azure platform** learned **configuration and deployment of models**.
- Worked on NLP tools **Semantic analysis, Nltk, Spacy,** Regex extraction (Python) for rule based, token search feature extraction.

Data science and ML

ACADEMIC PROJECTS

Sentiment analysis for Restaurant reviews using ML/NLP based model and learned NLTK, Stemming, Matplotlib, NumPy, Pandas, Python, Spyder, Scikit-learn.

Text identification using AI: Learned Backpropagation, Hyperparameter tuning, Matplotlib, Numpy, GIMP (customizing data), & developed from scratch each Python functions of Neural network for text detection on Jupyter.

Automatic Medicine Vending Machine Learned and utilized raspberry pi2, Image processing using OpenCV to detect object, colour, and count no. of medicinal tablets with contour detection.

AREA OF EXPERTISE

- NLP (Natural language processing) - Text mining, Topic modeling, sentiment & Semantic Similarity, N-gram, Word2Vec, TF-IDF, CBOW, Skip-gram
- Bert lang models, LSTM, classification, information extraction, document classif, Fraud detection
- Classific. Model - KNN, SVM, Logistic Regression
- Regression analysis
- CNN, object detection
- Jupyter notebook, GCP training, model optimization, parallelization, model evaluation.
- Deep learning framework - Tensorflow, Keras, Pytorch, Huggingface.

SKILLS

Prog lang and framework - Python, R basic, flask

Data visualization and manipulation – matplotlib, seaborn, plotly, tsne.

Model hosting - Azure, docker, git, GCP.

Python ML/NLP library - NumPy, regex, spacy, pandas, NLTK, scikit learn, zhon, Gensim.

ACHIEVEMENTS

Successfully completed Level 1, Level 2 Artificial Intelligence course in Intel.

First position in Algothon(best algorithm application) in Vellore Institute of Tech.

MTech Thesis paper on Clustering algorithm at Springer conference 2019, Spring Nature publications, A Survey on K-Means Clustering for Analyzing Variation in Data.

EDUCATION

Master of Technology

VIT university, Tamil Nadu

07/2017 - 05/2019, *CGPA:8.36/10*

Bachelor of engineering

Bharati Vidyapeeth, Mumbai

06/2012 - 06/2016, *CGPA:6.7/10*

Internship

Intel (software intern)

06/2018 - 06/2019