Preethi Patrick

Croydon, Greater London, UK · preethi.patrick5@gmail.com · +44 07918414376

in preethi-patrick-6803531b0 · ⌂ https://github.com/preethipatrick5

Summary

I am a passionate Data Scientist looking for an opportunity to grow as professionally and academically while contributing to the growth of an ambitious organization. I have 2 years of experience as an Assistant Professor in the Computer Science and Mathematics faculty. I had worked as a Data Science intern for an year at Knowsis. I have experience working on dataset cleaning, analysis to training advanced models such as BERT and other Deep Learning based models.

Education

University of Surrey, Guildford — Surrey
MSc Data Science — 2021 - 2023

Mohandas College of Engineering and Technology, Kerala University — Kerala, India

Master of Computer Applications *Percentage: 75* — 2014 - 2017

All Saints College, Kerala University — Kerala, India
BSc Mathematics *Percentage: 81* — 2011 - 2014

Professional Experience

Knowsis — London, United Kingdom
Data Science Intern — March 2022 - March 2023

- Worked with a team of two to develop and deploy models to analyze stock using twitter data.
- Was responsible for the development of sentiment analyzers using various ML techniques.
- Worked on text data using various NLP techniques for text mining.

Marian College of Arts and Science — Kerala
Assistant Professor — March 2018 - April 2020

- Held the role of Nodal Officer for National Scholarship.
- Managed the role of University Admission Coordinator.
- Mentored a small team of 5 junior python developers.
- Played key roles in the management of college website and student attendance systems.
- Responsible for planning exams and events.
- Effective communicator with excellent planning and organizational skills.

Projects

Trading & Investing using Twitter Data *BERT, Tensorflow, Python, NLP, Spacy, Transformers, Huggingface*

- Worked on dictionary creation for the model training.
- Worked on Data pre-processing on twitter dataset using lemmatization, stemming, stopword removal techniques.
- Worked on developing sentiment predictor using various models such as SVM and BERT.
- Worked on an inhouse model called AlphaVADER based on the VADER sentiment analyzer.

Deepfake Detection *Pytorch, Pytorch Lightning, Python, Xception, Resnet, FaceForensics, Deepfake Detection Challenge Dataset*

- Worked on deepfake detection as part of course dissertation.
- Used the FaceForensics & Deepfake Detection Challenge datasets for training & evaluations.
- Built models using Resnet, Xception & 3d Resnet to evaluate the performance of each model.
- Experimented with the robustness of each model by introducing random combinations of noices such as gaussian blur and rotation etc on the data.

- Retrained and evaluated the models on noisy data to observe the effects of retraining the models on noisy data.

Topic Modelling *BERT, Tensorflow, Python, TFS, Fast API, Luigi, AWS, Huggingface*

- Headed the data pre-processing and visualizations efforts.
- Worked on the model building, training and hyperparameter tuning.
- Setup ML pipline using luigi.
- Developed the API to interface with the TFS.

DB Design & BI *SQL, MS SQL Server Analysis Services*

- Reviewed the existing DB design supporting the online sales application.
- Designed and implemented new database design by applying normalization with optimisation of query execution time.
- Developed new stored procedures to deliver order processing functionality with error handling and transactional support to ensure that the data integrity.
- A POC BI platform using Microsoft SQL Server 2018 Analysis Services (SSAS/SSDT) was delivered for the given data requirement. Data Source View (DSV) and design corresponding to Cube measures and dimensions were created.

Evolutionary algorithms for optimisation and learning *Genetic Algorithm, Gradient descent, PSO, SL-PSO, NSGA II, Gray Coding, Efficient Non Dominated Sorting, Tournament Selection*

- Applied elitist non-dominated sorting genetic algorithm (NSGA-II) to solve multi-objective optimisation problem.
- Trained feed-forward multi- layer perceptron networks with two hidden layers to approximate the given function using memetic algorithms (Baldwin Learning , Lamarkian Learning).

ML - A comparative analysis on three dataset *Naive Bayes, MLP, DNN, DT, Gradient Boost, Random Forrest, SVM, K Means Clustering, ILP, Q-Learning*

- A comparative analysis is done using various metrics to identify which model works better for each dataset (Water Quality Dataset, Mushroom Classification Dataset & Pharma R&D Dataset).
- Q-Learning applied to solve Mountain Car Env from OpenAI Gym.

Customer Lifetime Value Prediction for E-commerce *R, DNN, K-means, RFM Analysis*

- Lead the efforts on Data Visualization & Statistical Analysis.
- Worked on RFM analysis on the dataset to generate CLV clusters.
- Developed DNN model on R for CLV cluster Prediction.
- Lead the team on model Visualization and Hyper parameter tuning.

Statistical analysis on Quiet Flows the Don book *Python, scipy, MLE, MAP*

- Used MLE to find the fit for Poisson distribution and Negative Binomial Distribution on the Quiet Flows the Don dataset which consists of sentence lengths of each sentence in the book.
- Used MAP to find the fit for a Poisson distribution with a Gaussian Prior.
- Analysed and compared the results of experiments.

Statistical analysis on ATLAS dataset *Python, scipy, MLE, MAP, FastAPI*

- Used Logistic regression and Naive Bayes models to predict if it is a background or signal class.
- Compare the models with cross validation and deploy using FastAPI.

PI Estimator *Python, boto3, AWS Lambda, AWS EC2, GAE, FastAPI*

- Developed a script that estimates pi based on the Monte Carlo approach.

- Developed an API to create n number of lambdas or EC2s as the user wishes and deploy the estimator code.
- Developed the UI where user can enter parameters such as number of simulations to be executed to estimate pi and the number of digits to be matched.

CloudVS *Ruby on Rails, Google Speech to Text API, Python*

- Developed a speech to text engine using google speech API
- Lead a team of 5 in the development of API using Rails.

CERTIFICATIONS

- Deep Learning Specialization, Deeplearning.AI
- NLP Specialization, Deeplearning.AI
- Convolutional Neural Networks, Deeplearning.AI
- Neural Networks & Deep Learning, Deeplearning.AI

ACHIEVEMENTS

- Bright Network Internship Experience - Worked on the Google Coding Challenge.