Ranjan Ravi

ranjan7796@gmail.com | LinkedIn | Portfolio | GitHub | (813) 729-1906 | Tampa, Florida

PROFESSIONAL SUMMARY

Experienced data scientist with a focus on Python and NLP, proficient in using NLP techniques for analysis, building NLP models, and data visualization. Adept at utilizing statistical methods and machine learning algorithms to extract insights and drive business decision-making. Passionate about solving real-world problems with data-driven solutions.

EDUCATION

Master's in Business Analytics and Information Systems, University of South Florida, Tampa, Florida Expected May 2023

Bachelor's in Computer Science and Engineering, Madras Institute of Technology, Anna University, India April 2018

WORK EXPERIENCE

Data Science Intern, Bayer, New Jersey June 2022 – Feb 2023

- Implemented Domain Adaptation using Generative Pseudo Labelling to fine-tune a pre-trained model with medical data. The process involved using GPL to create a training dataset from a large volume of unlabelled text data, which was then used to improve the model's performance through fine-tuning.
- Developed a semantic search tool allowing users to upload and access information from their corpus.
- Designed a custom ranking algorithm for search results, incorporating cosine similarity scores and fuzzy-matching distances. Wrote unit and integration tests for the API using Pytest.
- Assisted in deploying ML models (T5 and BERT) on Amazon Sagemaker.

Research Assistant and Tableau Mentor, University of South Florida, Tampa, Florida Nov 2021 – June 2022

- Involved in ongoing research of parsing student resumes using Named Entity Recognition.
- Developed a spacy annotator, a simple interface to quickly label entities for NER using ipywidgets which is used to generate training data with the spacy annotated data format.
- By working as a Teaching Assistant, I supported Undergraduate students in understanding the significance of data-driven decision-making and problem-solving by utilizing Tableau.

Data Scientist, TRIMBLE Inc, Chennai, India Nov 2020 – Aug 2021

- Built an efficient time series forecasting model for the Civil Construction division of Trimble, achieving 83% accuracy with the Facebook Prophet algorithm.
- Facilitated data management by collecting, cleaning, and loading revenue data from various sources into Snowflake using Python. Developed dashboards with monthly and quarterly revenue predictions using Tableau.
- Constructed an ML pipeline by deploying the model in AWS Sagemaker and implementing its inference function via AWS Lambda.

Associate Software Engineer, BNP PARIBAS India Solutions Private Limited, Chennai, India Jun 2018 – Nov 2020

- Involved in an Agile software design and development of Reporting platform for generating tax reports for clients.
- Written PL/SQL procedures and SQL scripts for implementing the tax calculation based on the Double Tax Treaty.

PROJECTS

- **NLP Albumentation:** Explored various data augmentation techniques for text data and evaluated the impact of NLP albumentation on the performance of a classification model. GitHub
- **Text Analytics:** Conducted text analysis on the inaugural speeches of all US presidents and performed Topic Modelling to uncover common themes in each speech. GitHub
- **Covid-19 Tracker**: Built a website with an interactive chart to display the trend of confirmed, recovered, and deceased coronavirus cases on a per-day basis using the Python Dash framework. GitHub
- **Sentiment Analyzer:** A big data project to capture the sentiment of E-Commerce user reviews stored in a Hadoop File System using Apache PySpark and MLlib. GitHub
- **GitHub Analytics:** Built a Classifier to identify potential candidates based on their GitHub profiles in order to minimize the time and effort the recruiters put in during the recruitment process. | GitHub
- **Linear Regression Model and Web Scrapping using Real Estate Data:** Designed and built a web scraper program to scrape house data from a real estate website and built a simple Linear Regression model | GitHub
- **Personal Portfolio:** Developed a compatible personal portfolio using HTML5/CSS3 and Bootstrap, and deployed it in the Heroku Cloud Platform. | URL | GitHub

TECHNICAL SKILLS

Programming Lang. & Frameworks:	Python, R, Flask, FastAPI, PL/SQL, Java, Dash, HTML & CSS
Databases:	Oracle, PostgreSQL, MS SQL, SQLite
Machine Learning Models:	NLP models, SVM, KNN, Ensembles, Neural Networks (CNN, RNN)
Cloud Technologies:	Amazon web services - EC2, S3, Sagemaker, Lambda, Snowflake, Azure ML Studio
Python Packages:	TensorFlow, PyTorch, Scikit-Learn, NLTK, SpaCy, Pandas, BeautifulSoup, OpenCV
Reporting Tools & packages:	Tableau, Matplotlib, Seaborn, Plotly, Jasper reports, PowerBI
Time Series Forecasting algorithms:	Facebook Prophet, Neural Prophet, ARIMA, Exponential Smoothing, LSTM