PAVITHRA DEVI M

pavipd495@gamil.com | Kaggle | LinkedIn | GitHub | 7904014309

EDUCATION

Manipal Institute of Technology **June 2023 - June 2025**
MSc Data Science *Grade: 9.8*
University College of Engineering (BIT campus) **June 2017 - June 2021**
Bachelor of Engineering in Computer Science *Grade:8.55*

TECHNICAL SKILLS

Programming Languages: Python, R, Java, SQL.
ML Frameworks: Keras, Scikit-learn, Pandas, Numpy, Tensorflow, AzureML, NLTK, Spacy.
DevOps: Docker, Kubernetes.
Professional tools: Github, Gitlab, Jira, Jenkins.
Data Base Management: MySQL, PostgreSQL, Vertica.
Web Development: HTML, CSS, Bootstrap, React-js, FastAPI.
NLP: Text-preprocessing, Language Modeling, NLTK, Spacy, LSTM, RNN, Word2Vec, Attention, Transformers, BERT.
Big Data technologies: Hadoop, Hive, Spark, Kafka.
Others: Luigi, Object Oriented Programming(OOPS), Shell Scripting.

PROJECTS

End-End Text Summarizer | Python, NLP, Transformers, Spaces, Docker, GitHub Actions, AWS, LLM Pull Request, GitHub
- **Fine tuned a LLM** model to summarize the given large documents using huggingface **transformers**.
- Analyzed the code-base and **fixed a production bug** in the huggingface transformers package during this project implementation.
- Deployed the fine-tuned model on Huggingface Spaces, creating an interactive web interface for users to input large documents and receive summaries.

EDA and Hypothesis testing | Python, p-value, T-test, Chi-squared test, ANOVA GitHub
- Conducted **EDA** on student grade performance dataset using Plotly and Seaborn to visualize data distributions and relationships.
- Performed **t-tests, chi-squared tests, and ANOVA**, to validate assumptions about variable interactions and their effects.

Next Word Predictor | Python, NLP, LSTM, Spacy, NLTK GitHub
- Engineered a **LSTM model** to predict the next word in sentences related to Medium article topics, achieving a **66%** accuracy.
- Implemented a pipeline encompassing **text pre-processing** and tokenization techniques to optimize model performance.

Disaster-tweet classifier | Python, NLP, Spacy, NLTK, Word2Vec GitHub
- Formulated a DL model with LSTM achieving **78% accuracy** to classify disaster tweets.
- Enhanced model performance through data preprocessing (**tokenization, lemmatization**) and utilization of **Word2Vec** vectors.

Machine learning models: Built from **Scratch** using Python Kaggle, GitHub
- Formulated Machine Learning models from scratch using Python, matching **90-99% accuracy to scikit-learn models**.

Models included: 1) Linear Regression, 2) Polynomial and Polynomial Ridge Regression, 3) Logistic Regression, 4) ElasticNet Regression, 5) Lasso and Ridge Regression, 6) Random Forest Regression, 7) Kmeans Clustering, 8) Decision Tree, 9) KNN Classifier.

Covid-19 Dashboard with Place-based Sentiment Analysis | Python, Dash, IBM Watson, scikit-learn. GitHub
- Implemented a live user interactive web dashboard that dynamically collects COVID-19 information from Government websites.
- Achieved a sentiment prediction model accuracy of **98.5%** and integrated CovidBot using **IBM Watson** to answer basic questions.

Automated Examination Management System (AEMS) | Python, OCR, IBM watson, scikit-learn, React-js, MongoDB GitHub
- Created a web app for document upload, implementing **OCR** for verification. Hall tickets issued within **10 minutes** if candidate eligibility is confirmed.
- Automated examination registration and hall ticket collection process, **saving human time up to 80%.**

WORK EXPERIENCE

Sandvine Technologies | Python, SQL, Docker, Kubernetes, Luigi, Kafka, React-js, FastAPI **May 2021 - Present**
Software Engineer II - Insights Data Storage (IDS) *Bengaluru*
Kafka SASL Authentication**:**
- Coded **SASL authentication** functionality for Kafka export, along with the validation of customer Kafka configurations.
- Enabled **Kafka** exports to broker with or without **SSL encryption**, ensuring seamless compatibility across **CNF & VNF.**

IDS Dashboard | React-js, SQLAlchemy, FastAPI, Bash scripting, celery
- Worked on backend and frontend components for IDS dashboard features, ensuring seamless integration and functionality.
- Delivering a solution for cluster activities, including monitoring and simplified CSV/GPG file uploads for TUI operations.

External Task Support:
- Developed this feature, enabling other **10+** Sandvine product teams to add custom tables and associated rollup tasks into IDS.

Product Codebase Improvements:
- Created **command-line functionality** to dynamically display migration task statuses, enhancing monitoring of migration process.
- Promptly addressed critical customer-reported bugs and provided expert assistance to resolve product-related inquiries.

Technocolaps | *Python, scikit-learn* **July 2020 - August 2020**
Machine learning Inter
- Used clustering and **feature selection techniques** to achieve **95%** accuracy in reducing traffic mortality rates.
- Deployed a Flask-based web application, providing users with intuitive access to valuable insights and functionalities.

The Spark Foundation | Python, scikit-learn, Flask **December 2020 - January 2021**
Machine learning Inter
- Developed a stock market prediction model using Keras with a dataset of **1M** data rows, achieving accurate forecasts. Deployed a web application to implement the model, providing users with access to stock market insights and predictions.

OTHER ACHIEVEMENTS

Kaggle Notebook Master for expertise in data analysis and machine learning.
Microsoft Azure ML scholar, Udacity & Microsoft 2020.
Fellow@Alpha.ai - specializing in Data Visualization, Data Storytelling.