

Pavan Kalaganda

(+91) 9676473214 | 142301014@smail.iitpkd.ac.in |  pavan-kalaganda |  pavan-kalaganda-07b28b288
 pavan-kalaganda.github.io/portfolio

Education

Indian Institute of Technology, Palakkad
Bachelor of Technology in Data Science and Engineering

Aug 2023 – Present
Expected May 2027

CGPA: 6.0/10

Relevant Coursework: Machine Learning, Deep Learning, Natural Language Processing, Artificial Intelligence, Probability & Statistics, Information Retrieval, Data Structures & Algorithms, Big Data, Database Management Systems

Technical Skills

Languages: Python, SQL, C/C++

ML & Deep Learning: PyTorch, Scikit-learn, Transformers (BERT, HuggingFace), Linear/Ridge Regression, PCA, K-Means, Time-Series Forecasting, Anomaly Detection

NLP & LLMs: Text Processing, BIO Tagging, Embedding-based Search, BM25, Learning-to-Rank, RAG, LLM Integration, Search Evaluation (NDCG, MRR, Precision, Recall)

Data Engineering: ETL Pipelines, Feature Engineering, Data Cleaning, Elasticsearch, API Integration

Databases: PostgreSQL, MongoDB, Query Optimization, Indexing, Window Functions

Visualization & Tools: Plotly, Matplotlib, Seaborn, Streamlit, Power BI (Basic), Git/GitHub, Linux, Jupyter Notebook

Projects

Nifty 50 Stock Market Prediction — *Python, Scikit-learn, Ridge Regression, PCA, K-Means*

- Engineered time-series features and built an end-to-end ML pipeline to forecast next-day Nifty 50 stock prices.
- Achieved $R^2 = 0.98$ with Ridge Regression after evaluating multiple regression models.
- Applied PCA and K-Means clustering to segment market regimes and surface trend patterns.

F1 Data Analytics & ML Dashboard — *Python, Scikit-learn, PCA, Streamlit*

- Processed 50,000+ race records and engineered ML features for driver and team performance analysis.
- Trained regression and clustering models achieving $R^2 = 0.90$ in lap-time prediction.
- Developed an interactive Streamlit dashboard that reduced manual analysis effort by 60%.

Numerical Search Engine over Wikipedia — *Python, Elasticsearch, BERT, HuggingFace, Wikidata*

- Built a numerical information retrieval system leveraging BIO tagging and Transformer-based models.
- Implemented hybrid BM25 + BERT reranking pipeline via Elasticsearch, improving retrieval relevance.
- Enabled structured numerical queries through Wikipedia-Wikidata entity linking.

Song Genre Prediction — *Python, Scikit-learn, Logistic Regression, Random Forest*

- Developed a genre classification model using audio features (tempo, energy, loudness) from song metadata.
- Applied feature engineering and preprocessing; evaluated with accuracy and macro F1-score metrics.
- Compared Logistic Regression and Random Forest classifiers to select optimal model.

Achievements

- Selected to represent IIT Palakkad at the Inter-IIT Weightlifting Competition, competing at a national-level platform.
- Captained IIT Palakkad Powerlifting Team to a Gold Medal at the College General Championship 2026.