Kola Sai Deepthi Priya

Hyderabad, India | +91-9110330726 | deepthikola40@gmail.com | LinkedIn | GitHub

---

Professional Summary

Detail-oriented Data Science professional with 5.5 years of IT experience, including 3.5 years in AI/ML-focused roles. Specialized in supporting the full AI lifecycle through **data extraction, annotation, exploratory data analysis (EDA)**, and **model evaluation workflows**. Skilled in validating model performance, augmenting datasets to improve generalization, and ensuring data integrity across structured and unstructured sources. Aware of applying **ML and NLP techniques** (including BERT) to classification and prediction tasks. Adept at transforming raw data into actionable insights that support high-impact AI solutions.

Skills

- **Languages:** Python, SQL, Java
- **ML/Statistics:** PCA, VIF, RFE, Linear/Logistic Regression, Decision Trees, Random Forest, K-means, Hierarchical Clustering
- **Libraries/Packages:** Scikit-Learn, NumPy, SciPy, Pandas, Seaborn, Matplotlib, SpaCy, NLTK, Transformers (BERT)
- **NLP & Text Analytics:** Named Entity Recognition (NER), Text Classification, Tokenization, POS Tagging, Text Preprocessing
- **Other Tools:** Data Visualization (Python), Data Annotation, Data Extraction, Data Analysis (Excel), Data Validation

---

Professional Experience

Data Scientist / Lead Engineer | ACL Digital & HCL Technologies (PayPal Project - CW) *Jul 2021 – Present* | *Hyderabad, India (Transitioned from ACL Digital to HCL Technologies in May 2023 on the same client project)*

- Developed and deployed **Machine Learning and NLP models** for structured and unstructured privacy data classification, enhancing data governance.
- Implemented **BERT for Named Entity Recognition (NER)** using PyTorch; annotated data with regex and custom logic to improve model accuracy.
- Conducted extensive **Exploratory Data Analysis (EDA)**, automated model evaluation, and meticulously validated True Positive/False Positive (TP/FP) outputs.
- Applied **data augmentation techniques** to significantly improve model training and classification performance.

- Collaborated closely with Subject Matter Experts (SMEs) to establish logical data mappings, ensuring high data quality and integrity.
- Delivered comprehensive technical solutions, maintained thorough documentation, and provided regular, impactful stakeholder updates.

Programmer Analyst | Cognizant Technology Solutions *Sep 2018 – Apr 2020* | *Chennai, India*

- Integrated and harmonized data from **over 20 disparate systems** to facilitate healthcare and insurance analytics.
- Resolved critical data quality issues and implemented robust data validation frameworks.
- Developed insightful dashboards and collaborated with analysts for efficient data transformation processes.

Data Analyst Intern | The Sparks Foundation *Mar 2021 – Apr 2021* | *Bangalore, India*

- Built and evaluated **regression and clustering models** on various datasets, demonstrating proficiency in data cleaning, validation, and visual analytics.

Education

M.Sc. in Data Science | Liverpool John Moores University, UK *Apr 2025 – Apr 2026 (Expected)*

PG Diploma in Data Science | IIIT Bengaluru *Feb 2020 – Mar 2021*

B. Tech in Computer Science and Systems Engineering | Sree Vidyanikethan Engineering College, Tirupati *Aug 2014 – Jun 2018*

Additional Projects

Bike Price Prediction | *Predictive Analytics | Python, Jupyter Notebook*

- Developed a **linear regression model** using VIF for feature selection to accurately predict American bike prices. Achieved an impressive **90.5% R^2** on test data using 6 optimal features, demonstrating strong predictive capability.

Lead Scoring | *Predictive Analytics | Python, Jupyter Notebook*

- Classified potential sales leads using **logistic regression and Recursive Feature Elimination (RFE)**. Attained an **AUC of 0.88**, **Sensitivity of 0.80**, and an **Accuracy of 0.79**, enabling targeted outreach.

Telecom Churn Analysis | *Predictive Analytics | Python, Jupyter Notebook*

- Identified high-churn risk customers for a telecom firm utilizing **logistic regression and random forests**. Achieved an **AUC of 0.94** and an **F1-score of 0.68**, providing actionable insights for customer retention strategies.