

# Sindura Reddy Challa

Maryland, USA | [sindurachalla@gmail.com](mailto:sindurachalla@gmail.com) | [Sindura-portfolio](#) | [linkedin.com/in/challasindurareddy](https://linkedin.com/in/challasindurareddy) | [github.com/Sindura06](https://github.com/Sindura06) | +1 (571)246-1039

## EDUCATION

**University of Maryland, Baltimore County**  
*MPS in Data Science*

Maryland, USA  
Aug 2023 – May 2025

**Vidya Jyothi Institute of Technology**  
*B.Tech in Information Technology*

Hyderabad, India  
Aug 2019 - May 2023

## TECHNICAL SKILLS

**Languages & Libraries:** Python (Pandas, NumPy, Scikit-learn, TensorFlow), SQL

**Machine Learning:** Supervised/Unsupervised Learning, Time Series Forecasting, Classification, Regression, XGBoost, Random Forest, LSTM, Feature Engineering, Model Evaluation (MAE, RMSE, R<sup>2</sup>)

**NLP:** Transformers (BERT, DistilBERT, Pegasus, BART), Zero-Shot Classification, Summarization

**Data Analytics:** Statistical Analysis, Hypothesis Testing, Data Cleaning, Web Scraping, ETL Pipelines

**Visualization & BI Tools:** Tableau, Matplotlib, Seaborn, Plotly

**Deployment Tools:** Streamlit, FastAPI, GitHub, Databricks, Agile

## EXPERIENCE

**Data Analyst Intern**

Apr 2021 - Jun 2021

*Knowledge Solutions India*

*Remote*

- Conducted exploratory data analysis and preprocessing to uncover trends in patient health records, leading to the development of risk stratification models with 87% classification accuracy and 0.91 AUC.
- Enhanced diagnostic accuracy by reducing false negatives by 23% and improving recall by 18%, achieved through feature engineering, correlation analysis, and rigorous model evaluation.
- Designed interactive dashboards with 12+ visualizations (heatmaps, ROC curves, feature rankings) using SQL and Python, enabling stakeholders to make informed decisions on healthcare policies.

## PROJECTS

**Highway Maintenance Forecasting Model**

Apr 2025

- Predicted long-term pavement degradation with 30% improvement in accuracy (lower RMSE) by integrating 10 years of traffic and surface condition data into a unified geospatial model combining HPMS and FAF records.
- Boosted model performance through temporal feature engineering and advanced algorithms, including ensemble methods (Random Forest, XGBoost) and sequence modeling with LSTM networks.
- Developed interactive geospatial dashboards to visualize deterioration patterns and identify high-risk segments across 2,000+ highway segments, supporting proactive infrastructure planning simulations.

**Research Paper Classification & Summarization**

Dec 2024

- Achieved 92% classification accuracy across 170+ scientific fields by fine-tuning DistilBERT for fast, real-time document categorization, balancing performance with computational efficiency.
- Generated concise, domain-specific summaries with an 83% ROUGE-L score by integrating Pegasus, a model optimized for long-form abstractive summarization, outperforming alternatives like T5 and BART.
- Built a seamless PDF-to-summary pipeline by developing a Streamlit app with backend classification and summarization, allowing users to upload research papers and extract key insights instantly.

**Mood-Based Book Recommendation System**

Sep 2024

- Enabled personalized, emotion-aware book discovery across 8 mood categories by leveraging the BART NLI model for zero-shot classification, moving beyond traditional genre-based filtering.
- Built a complete data pipeline that processed 1,200+ book entries with rich metadata, using automated web scraping and premise-hypothesis inference to assign mood labels accurately.
- Deployed a real-time recommendation system through a Streamlit frontend and FastAPI backend, allowing users to receive dynamic suggestions based on mood classification.