

# PRERANA HADADI

+1(720) 312-3046 | Boulder, CO | [prerana.hadadi@gmail.com](mailto:prerana.hadadi@gmail.com) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

**Master of Data Science** **Aug 2023 - May 2025**  
University of Colorado Boulder GPA: 3.98/4.00  
**Relevant Coursework:** Data Mining, Machine Learning, Statistical Methods and Applications I & II, Cyber Security for Data Science

**Bachelor of Computer Science** **Aug 2019 - May 2023**  
PES University GPA: 8.23/10.00

## EXPERIENCE

**Data Scientist Intern** **Jan 2023 - Jun 2023**  
Textron *Bengaluru, IN*

- Architected a machine learning model on **Databricks** and **Azure Cloud** to improve network security, reducing threat response time by **40%** and decreasing data processing duration by **70%**.
- Constructed a **K-means** clustering model with **Cross-Validation**, tested on **200+** compromised websites ; increased model fairness by **4%**, enhancing threat detection and minimizing false positives.

**Machine Learning Intern** **May 2022 - Jul 2022**  
Center for Data Science and Machine Learning - PES University *Bengaluru, IN*

- Managed a high-impact **Machine Learning** research project examining heart disease, discovering **8 insights** that enhanced predictive models and elevated patient outcome success rates by **35%**.
- Led testing protocols, revealing the Random Forest algorithm's top performance with a **91.48%** correctness rate, thus improving the project's overall predictive exactness.

## PROJECTS

**News Insight Analyzer ([Github Link](#))** **Jan 2024 - May 2024**

- Enhanced information retrieval by processing 1,200+ URLs and extracting article content using **LangChain**, achieving a **92%** accuracy rate in retrieval with **OpenAI embeddings** and **FAISS**.
- Optimized LLM interaction by developing a system for seamless LLM queries, processing 600+ queries, increasing user satisfaction by **30%**, and reducing response time by **25%**.

**YouTube Title Ranking System for Content Creators ([Github Link](#))** **Jan 2024 - May 2024**

- Innovated a predictive model leveraging **BERT** for title classification and **Elastic Net Regression** for view forecasting by boosting content creators' engagement by **25%** through video title optimization strategies.
- Implemented **advanced data collection** techniques, including **web scraping** and **YouTube API**, to gather and analyze data from over **50,000** YouTube videos across genres, enhancing the system's reliability.

**Automated Regional Language Subtitles for English Videos ([Github Link](#))** **Jan 2022 - Jan 2023**

- Pioneered a sophisticated speech recognition system using **Python, Vosk, Spleeter, FFmpeg and Tkinter**; achieved a **35%** improvement in user accessibility by adding perfectly synchronized subtitles and enhancing audio clarity.
- Accomplished an **85%** precision rate for American accent detection by integrating state-of-the-art testing metrics, resulting in a **25%** increase in user satisfaction.

**Sentiment Analysis using Pyspark ([Github Link](#))** **Aug 2022 - Dec 2022**

- Designed an advanced **Big Data** infrastructure using **Twitter API and Kafka**, capturing and transforming live tweets into RDD format, which improved analytical processing speed by **60%** and provided actionable insights.
- Directed comprehensive training initiatives for Machine Learning models using batch sizes of **20,000**; attained **87%** with Bernoulli NB, boosting predictive analytics with SGDClassifier and PassiveAggressive Classifier.

## SKILLS

<b>Languages</b>	: Python (NumPy, Pandas, Scikit-learn, Matplotlib, TensorFlow, Keras), SQL, R, C, MATLAB, Pytorch
<b>Frameworks</b>	: Langchain, Databricks ,Open CV, Spark, Kafka, Excel, Hadoop, Azure Cloud, Regression, Classification, Clustering, PCA, Data Mining, Tableau ,PowerBI
<b>Data Science Technologies</b>	: A/B Testing, Data science pipeline (cleaning, wrangling, visualization, modeling, interpretation)
<b>Certifications</b>	: Neural Networks and Deep Learning, Data Analysis with Tidyverse, Blockchain
<b>Publications</b>	: "A novel approach in the automatic generation of regional language subtitles for videos in English", IEEE 2023 April