Anudeep Nayak

github.com/Anudeep-Nayak | 303-269-9785 | anna5868@colorado.edu | Linkedin.com/Anudeep-Nayak/

EDUCATION

Master's in Data Science *August 2023 - Present*
University of Colorado Boulder, Boulder, CO 3.96/4.0

Bachelor of Technology in Computer Science *June 2018 - July 2022*
Manipal Institute of Technology, Manipal, India 3.39/4.0

TECHNICAL SKILLS

- **Languages/Libraries:** Python, Pandas, R, Oracle SQL, PySpark, MongoDB, Scala, Numpy, TensorFlow, Keras, Matplotlib, ggplot2, Dplyr, Scikit-learn, Seaborn, OpenCV, Spacy, NLTK, Genim
- **Data Science**: EDA, Data Cleaning & Mining, Data Visualization, Unsupervised and Supervised Machine Learning models, Clustering, Time Series Analysis, Statistical Analysis, NLP, Regression, Regularization, ML Pipelines, Recommendation Systems, LLMs, XGBoost
- **ETL Tools/Techniques:** ETL Pipelines, Data Warehousing, Big Data Pipelines, Data Engineering,, AWS Lambda, Autosys, Git CI/CD process, Azure Synapse, Hadoop, Apache Spark
- **Developer Tools:** Git, Microsoft Office, AWS, Azure, Tableau, PowerBI, APIs endpoints, Docker

EXPERIENCE

Software Engineer *January 2022 – June 2022*
GE Healthcare Bangalore, India

- Utilizing Java, and SQL introduced new features and enhancements biweekly in agile development.
- Resolved Data inconsistency and defects, collectively bringing down defect rate by 80%.

Software Engineer Intern *July 2022 – July 2023*
GE Renewables Bangalore, India

- Built a model that categorizes incoming inconsistent data into right categories using fuzzy logic with 98% accuracy.
- Designed and implemented NET Promoter Score (NPS) component to KPI of the system.

PROJECTS

Resume Filter, GE Renewable Energy

- Secured first place in Intern's Hackathon for developing a model that filters resume into different roles.
- With limited training data present, bootstrapping and cross validation techniques were employed to train and validate the model.
- Implemented KNN model to train the data, accuracy of 83% was observed in test set.

COVID-19 Analysis, Coursera (University of Colorado Boulder)

- COVID-19 data is taken from New York times to analyze how it has impacted USA through case and death statistics.
- Tidyverse and R is used to compare the statistics across different states in USA.

Cinecypher, University of Colorado Boulder

- Built a recommendation system of movies based on description or storyline given.
- Data is webscraped through imdb website and cleaned using pandas and regex.
- Implemented NLP techniques LSI(Latent Semantic Index) and tf-idf on text data using spacy and genim.

Audio Emotion Analysis, University of Colorado Boulder

- Audio is extracted and converted to text from movie using openAI api.
- Feedforward neural network is used to classify the emotions from the text data, textBlob is used to classify the sentiments. Data analysis is then conducted to see various Emotional and sentimental aspects of the movie.
- Correlation analysis revealed no correlation between emotions and sentiments.

Movie Runtime vs Commercial Success, University of Colorado Boulder

- In R, a hypothesis test was performed to assess the association between longer movie runtimes and higher ratings and revenue.
- Z test conducted on the data resulted in p value close to 1%, proving the hypothesis.