Riteesh Ram

Columbia, MO
(908) 447-2924 / riteeshram.22@gmail.com
lwww.linkedin.com/in/riteesh-ram2215

Education

University of Missouri, Columbia. Expected May 2023 ***Master's in Data Science & Analytics*** **(SGPA - 4.0)**

KLE Technological University Hubli . June 2021 ***Bachelors in Computer Science and Engineering*****(CGPA - 3.6)**

Experience

Graduate Student Assistant Jan 2022 - Present, Position – Graduate Student Assistant **Advisor** - Mihail Popescu

• Part of the Image-guided Biocuration of disease pathway from scientific literature project. All the articles which contain the gene pathway images and texts together make up the dataset. A combination of "OCR" and the "RetinaNet" model is used to detect the gene names and the gene interactions.

Skills

• **Programming:** SQL**,** Python (NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, Geopandas), R (dpylr, Tidyverse, ggplot2) • **Tools:** PostgreSQL, Jupyter Notebook, Colaboratory, Tableau, Visual Studio, RStudio, PyCharm, Excel, PowerPoint

• **Domain Knowledge:** NLP, EDA, ETL, ERD, Normalization, Feature Selection, Anomaly Detection, Dimensional Reduction, Statistics, Model Building, Machine Learning Pipelines, Model Development, Model Validation, Visualization, Data Story Telling, Business Analytics

• **Relevant Courses:** Python for Data Science, Database and Analytics, Big Data Visualization, Statistical & Mathematical Foundation, Machine Learning, Big Data Security, Data Analytics from AML, Tableau

Projects

Extractive text Summarization:

- Created a Website which takes user input as text or document.
- Implemented BERT model which which is used to summarize an input which can either be a document or text containing 1000+ words and to a meaningful text containing 300 words

Multiple Object tracking in 360 degree videos:

- Designed a machine learning model which can detect and track objects still or moving in a 360 degree video.
- Yolo 360 was used to detect the objects present in the video
- Hungarian algorithm used to track the movement of the objects based on the bounding boxes which was produced by the Yolo 360 model.

Image-guided Biocuration of disease pathway from scientific literature project:

- Created a website which can identify gene names and can link the genes together based on the pathway images and biological articles available.
- Bio bert model is used to identify the gene names from the articles and a combination of OCR and RetinaNET is used to detect the gene names from the images.

NAEP data mining competetion:

- Patricipated in a competition hosted by kaggle which consisted of a student exam info.
- The task was to predict the efficiency of the student who gives a math test.
- Various data preprocessing techniques used to clean the dataset which consisted of time variables, predicted the result using XGBoost model which had the highest accuracy of all the models

Loan Prediction and analysis:

- In this project the task was to judge whether the customer is eligible for loan or not.
- The data needed huge amount of pre-processing work since the number of missing values and outliers were more.
- Out of all the machine learning models, random forest produced better accuracy.

Netflix data Visualization :

- Project mainly based on the Exploratory data analysis part.
- Created all graphs using ggplot which gives a clear idea about the dataset.
- Also handled all the missing values and outliers.

Publications :

- Published a research paper on "Extractive Text Summarization using BERT".
- https://link.springer.com/chapter/10.1007/978-981-16-6407-6_63