Rushikesh P Dhumal

r.dhumal@rutgers.edu | (732) 685-0337 | linkedin.com/in/rushikeshdhumal/ | rushikeshdhumal.github.io

Aspiring data scientist with 2+ years of professional and internship experience applying advanced analytical techniques to solve complex business problems. Proficient in Python, R, SQL, and machine learning frameworks like TensorFlow and PyTorch. Skilled in building ETL pipelines, automating workflows, and developing predictive models that drive data-informed decisions. Experienced in data visualization with Tableau and Power BI to deliver actionable insights. Proven track record of leading data-driven projects, including building NLP models and optimizing database systems. I am pursuing an MS in Data Science at Rutgers University and am eager to leverage my skills to contribute to impactful solutions in a dynamic workplace.

Education

Master of Science in Data Science, Rutgers University, New Brunswick (GPA: 3.67/4) **August 2024 — May 2026**

Technical Skills

Programming/Scripting	Python, R, PowerShell, Git, SAS, MATLAB, GNU Octave, C++, C, JavaScript, HTML5, CSS3
Machine Learning & AI	Scikit-learn, TensorFlow, PyTorch, Keras, AutoML, OpenCV, NLP (spaCy, NLTK), LLMs
Data Engineering & ETL	SQL, NoSQL, MongoDB, Databricks, Apache Kafka, Airflow, Snowflake
Cloud & Big Data	Hadoop, Spark, AWS RDS, AWS S3, AWS SageMaker, AWS Bedrock, Azure ML, Google Cloud AI
Visualization & BI	Power BI, Tableau, Looker Studio, Oracle NetSuite, MS Excel, Seaborn, Matplotlib, Plotly
Automation, Webscraping	Selenium, BeautifulSoup, API Integration, Web Crawlers

Professional Experience

Fields Data - Data Analyst Consultant **July 2022 — July 2024**

- Developed an NLP-driven recommendation engine using Python (spaCy, NLTK) to match 600+ humanitarian organizations with relevant projects, improving partner identification efficiency.
- Automated data ingestion pipelines for 300+ datasets, leveraging Python APIs, reducing manual workload by 80%.
- Optimized database performance on AWS Cloud RDS, reducing query latency by 40% by indexing and schema optimization.
- Designed and deployed 'Data Terns', a data service marketplace, using full-stack development principles and cloud technologies.

Internship Experience

Nestle - Sales Analytics Intern **March 2022 — May 2022**

- Analyzed retail data from 1600+ outlets to identify underperforming regions and refine sales forecasts, driving strategic decision-making.
- Enhanced demand forecasting models, leading to improved customer segmentation and resource allocation.

SciSpace - Data Mining Intern **March 2022 — May 2022**

- Built a scalable data extraction pipeline using Python & web scraping tools, improving lead generation by 10%.
- Optimized data retrieval processes, integrating automation tools like Mailscoop.io, Skrapp.io, and Hunter.io.

Fields Data - Data Analyst Intern **April 2021 — May 2022**

- Automated data management pipelines using Python, APIs, and web scraping (BeautifulSoup), streamlining data consolidation and reporting.
- Developed interactive dashboards in Power BI & Google Looker Studio, enhancing data-driven decision-making.
- Delivered weekly technical training sessions on machine learning concepts, including neural networks and decision trees.

Projects undertaken

Organisational project opportunity matching (Fields Data)

- Developed a recommendation engine to match organizations with project opportunities using Natural Language Processing (NLP) techniques in Python (spaCy, NLTK, scikit-learn).
- Extracted and preprocessed unstructured text from project descriptions and organization profiles using TF-IDF, word embeddings (Word2Vec, GloVe), and Named Entity Recognition (NER).
- Built a similarity ranking algorithm with cosine similarity, topic modeling (Latent Dirichlet Allocation – LDA), and deep learning-based embeddings.

SpaceX Falcon9 First Stage Landing Prediction (Github link)

- Developed a classification model to predict successful Falcon9 first-stage landings using launch data scraped from SpaceX.
- Performed feature engineering and trained and evaluated predictive models including Logistic Regression, Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors (KNN).
- Improved model accuracy by 20% through hyperparameter tuning (GridSearchCV, RandomizedSearchCV) and ensemble learning techniques (Bagging, Boosting).
- Created a dynamic interactive dashboard using Plotly Dash, enabling real-time exploration of launch outcomes, payload performance, and success probability distributions.