Ravi Kumar

LinkedIn | GitHub | Medium

Email : raviikrds@gmail.com

EXPERIENCE

CAPGEMINI — *Apr 2021 – Present*

Associate Consultant (Data Scientist) *Dec 2023 – Present*

- Built scalable data pipelines & ML pipelines for text analytics, automating data ingestions, processing and model deployment workflows.
- Designed a text classification solution to categorize ServiceNow tickets, coupled with real-time Power BI dashboards for stakeholder visibility.
- Developed end-to-end ML models, Improving accuracy from 55% to 83% through advanced feature engineering, hyperparameter tuning and continuous enhancement.
- Developed the Email Classifier framework to fetch emails, classify content, trigger ticket creation in ServiceNow, and auto-assign tasks to respective teams.
- Proactively enhanced the product by debugging, identifying new features, and aligning ML capabilities with business growth opportunities.
- Collaborated with cross-functional teams (engineering, product & business) to optimize ML models and enhance products features.

Senior Software Engineer (Data Scientist) *June 2022 - Dec 2023*

- Developed and deployed end to end NLP-based text classification models using Python, Flask, and ML frameworks.
- Handled data cleaning, data preprocessing, feature extraction, and pipeline automation using Pandas and NLTK.
- Worked on deploying AI solutions on AWS and On Prem server ensuring scalability and performance.
- Contributed to internal POCs and tools to support ML adoption, while mentoring junior team members on best practices in NLP and deployment.

Software Engineer (Data Analyst) *Apr 2021 - June 2022*

- Extracting and transforming data from Amazon Redshift using SQL and Python to support end to end workflows.
- Developing and maintaining Power BI dashboards to visualize outputs, trends, and key performance indicators for stakeholders.
- Led dashboard migration from Tableau to Power BI, resulting in improved performance and user experience.
- Implementing dashboard enhancements based on client requirements, including new KPIs, filters, and drill-down functionalities.
- Worked with cross-functional teams to ensure accurate, consistent, and actionable data reporting.

PROJECTS

Twitter Sentiment Analysis with end-to-end AWS Data & ML pipeline *Jul 2023 - Jan 2024*

- Designed an automated NLP pipeline to ingest tweets via Tweepy, process them with PySpark, and store results in Delta Lake for scalable analysis.
- Trained sentiment models (Logistic Regression, BERT) using Scikit-learn and Hugging Face Transformers, achieving 85%+ accuracy on labeled tweet datasets.
- Used Apache Airflow for end-to-end workflow orchestration and MLflow to track experiments, metrics, and model versions.
- Deployed a Streamlit dashboard to visualize real-time sentiment trends for social insights and engagement analysis.

AI-based Auto-Tagging for Legal or Research Documents *Jul 2022 - Sept 2022*

- Developed an NLP pipeline to auto-summarize and tag legal/research documents (6–10 pages) in English and Hindi using pretrained models like BART, mT5, and IndicBAR)
- Extracted contextual tags using BERTopic and KeyBERT with multilingual sentence embeddings for accurate classification.
- Integrated PDF parsing, language detection, and MySQL for end-to-end document processing and storage.
- Reduced manual document review time by 70% and enabled faster content discovery for legal and academic users.

AI-Powered Deepfake Video Detection System *Jul 2022 - Sept 2022*

- Developed an AI-powered deepfake detection tool using XceptionNet and OpenCV to identify manipulated facial frames in videos.
- Built an end-to-end pipeline with frame extraction, face alignment, and real/fake classification with accuracy metrics.
- Integrated a lightweight Streamlit dashboard to upload videos and visualize prediction results with confidence scores
- Demonstrated robust detection across FaceForensics++ and DFDC datasets, focusing on both visual and audio inconsistencies.

EDUCATION

LAKSHMI NARAIN COLLEGE OF TECHNOLOGY — *Bhopal, Madhya Pradesh*

- B.E. in Electronics and Communication Engineering *June 2016 - June 2020*
  CGPA: 8.19/10

SKILLS

Programming Language: Python, SQL, DSA
Programming Frameworks: Pandas, NumPy, Matplotlib, Seaborn. TensorFlow, SciKit-Learn, Flask
Technologies: Deep Learning, NLP, Docker, Git, AWS SageMaker, Machine Learning, Microsoft Power BI, AI, Databricks
Soft Skills: Problem Solving, Storytelling with Data, Presentation Skills, Collaboration & Teamwork, Leadership, Mentorship

ACCOMPLISHMENTS

Key Accomplishments

- Conducted a virtual webinar on "Storytelling with Data" for 1000+ colleagues, enhancing data visualization and communication skills across teams.
- Published multiple blogs on Medium covering Data Science, Machine Learning, & AI contributing to the community and knowledge sharin.
- 2nd Place in Hackathon: Awarded 2nd place in a high-level hackathon focused on developing an Advanced Farming System, demonstrating technical innovation and team collaboration.
- 5☆ Python Rating on HackerRank: Achieved a 5-star rating in python highlighting strong programming skills and advanced algorithmic proficiency.
- 4☆ Coder Rating on HackerRank: Earned a 4-star rating on HackerRank, showcasing ability to solve complex coding challenges and algorithmic problems efficiently.