HOZEFA PATEL

(Big Data Analyst)

+1(226)-898-3229 | hozefapatel9@gmail.com | https://github.com/hoz-efa |
https://www.linkedin.com/in/hozefa-patel

ABOUT

I am a graduate in Big Data Architecture with practical experience in Python and R-programming along with some front-end technologies including - HTML/CSS, PHP and proficient in relational databases including MSSQL, PostgreSQL, MongoDB. Having hands-on experience in Intergrating API on pipeline using Kafka and implementing and training Spark ML model. Also, worked on wordpress to implement e-commerce and product-based applications and deployed on DevOps or Git.

EDUCATIONAL BACKGROUND

Artificial Intelligence with Machine Learning
[Jan 2024 - Present]
Humber College, Toronto, ON
[Skills: AI, Machine Learning Techniques, AWS, Image Processing, Natural Language Processing, Elastic Search, Kibana, Logstash]

Big Data Solution Architecture
[May 2023 - Dec 2023]
Conestog College, Kitchener, ON   [78.36 / 100]
[Skills: Python, Hadoop, Tableau, RDBMS, Kafka, Google Cloud Platform, R-programming]

Bachelor of Computer Applications
[2018 – 2021]
Saurashtra University, India   [71.30 / 100]

SKILLS AND PROFICIENCIES

Programming Languages: Python, R-programming, C/ C++, SQL.

Backend Technologies: R-programming (Machine Learning and Data Analysis), Django, PostgreSQL, MySQL, and MongoDB, Pyrogram, Python-Telegram-Bot, Heroku, Postman, REST API, MSSQL, Git, Azure DevOps, Google Cloud Platform, Hadoop, Apache Kafka (API consumption, producer, consumer, pipelines), Spark (dataframes, Machine Learning Models and Training, Data Analysis), Hive, RDBMS, Data Structures, OOPs.

Frontend Technologies: HTML5, CSS, Bootstrap, PHP, Wordpress.

Secondary Skills: Sellenium, Agile Scrum.

RECOGNITIONS AND AWARDS

- Secured 1st position at the 11th National Science Symposium in 2019.
- Achieved 2nd position with a team in hackathon at TechnoMark in 2022.
- Received a prestigious certificate from the Science and Engineering Research Board of the Department of Science and Technology, Government of India.
- Participated in 'Prakhara-Bodhan', a National level technical event in 2019, leading workshops on the 'Internet of Things' and 'Ethical Hacking', highlighting a proactive approach to understanding and applying critical aspects of technology and cybersecurity.

LANGUAGES KNOWN

English *(Advanced),* Hindi, Gujarati

HANDS-ON EXPERIENCE

Project: Enhancing Cybersecurity Resilience through Advanced Data Analysis and Risk Management

Description: Spearheaded a cybersecurity initiative aimed at developing a risk assessment model to enhance organizational cybersecurity measures. Leveraged machine learning models to analyze cybersecurity metrics and calculate risk scores, while providing actionable recommendations through a custom-built web portal.

Tools and Technologies: Python (Pandas, NumPy, Scikit-learn), Flask, SQLite, GEMMA-2-2b-it, HTML/CSS, JavaScript, Plotly.

Responsibilities:
- Conducted data cleaning, feature *standardization, and transformation* on simulated cybersecurity datasets using Python.
- Implemented multiple *machine learning models* (Random Forest, XGBoost, etc.) to calculate risk scores, evaluating model performance through MSE and R^2 metrics.
- Developed a web portal using *Flask and SQLAlchemy*, enabling both single and batch risk assessments with interactive data visualization.
- Integrated the *GEMMA-2-2bit model* to generate personalized, actionable security recommendations based on identified vulnerabilities.
- Collaborated with stakeholders to ensure *interactive visualizations and reports* met specific organizational needs, supporting strategic cybersecurity decisions.
- Addressed challenges with seamless data flow and optimized the web portal's client-server architecture for responsive, cross-device usability.
- Provided insights on future improvements, including expanding datasets to cover a wider range of industries and continuously fine-tuning the model.

Project: Analysis of News Articles using Apache Spark & Kafka

Description: This project focused on developing an end-to-end big data pipeline to analyze sentiment in news articles. The pipeline ingests data from NewsAPI, processes it using Apache Kafka and HDFS, and employs Spark ML for sentiment classification.

Tools and Technologies: Apache Spark, Scala DataFrame, Apache Kafka, HDFS, Spark ML, NewsAPI.

Responsibilities:
- *ETL Process Development:* Set up an efficient ETL process using Apache Kafka, ensuring continuous data flow from NewsAPI to HDFS.
- *Data Analysis with Spark SQL:* Utilized Scala DataFrame to perform insightful queries on the ingested data, deriving meaningful patterns and trends.
- *Machine Learning Model Implementation:* Trained a Spark ML classification algorithm to accurately predict the sentiment (positive or negative) of news articles, demonstrating a strong grasp of machine learning concepts and Spark ML libraries.
- *Technical Problem-Solving:* Overcame various technical challenges during setup, including configuring Kafka brokers and Hive integration, enhancing my problem-solving skills and technical proficiency.

Project: Business Intelligence Implementation & Artwork Data Analysis

Description: Led a comprehensive Business Intelligence initiative at the Waterloo Museum of Art, focusing on both preprocessing a large dataset of art collections and creating insightful visualizations.

Tools and Technologies: Python (Pandas, NumPy, Matplotlib, Seaborn), Tableau.

Responsibilities:
- *Data Cleaning and Validation* with Python.
- Transitioned from Python exploratory visuals to Tableau for more sophisticated and interactive *data visualization*.
- Crafted data-driven narratives that highlighted key findings from the museum's collection, aiding in curatorial decisions and exhibition planning.
- Provided museum administrators and stakeholders with actionable insights derived from the visualizations, supporting strategic decisions related to art acquisitions and displays.
- Addressed various technical challenges during the project, such as *integrating Python scripts with Tableau* and ensuring seamless data flow.
- Collaborated closely with museum staff to tailor visualizations that met their specific