

MANAV CHETWANI

<https://www.linkedin.com/in/manavchetwani/>

Toronto, ON
(437) 599-4919
manav01chetwani@gmail.com
<https://github.com/mnx02>

SUMMARY

- Over 2+ years of work experience as a Machine Learning Engineer and Data Engineer.
- Skilled in architecting end-to-end ETL processes, harnessing technologies such as **Spark**, **Spark Streaming**, **Databricks**, and **Azure Data Factory**, to efficiently manage and process **large-scale data pipelines**.
- Proficient in developing and deploying machine learning models, specialized in **DNN** and **XGBoost**, to analyze and mitigate anomaly behaviors within large datasets, enhancing security infrastructure.
- Strong coding abilities in **Python** for efficient data manipulation and transformation.

TECHNICAL SKILLS

Scripting Languages	C, C++, Python, Java, Scala, HTML, Shell Scripting
Big Data Processing Tools	Apache Hadoop, Apache Spark, Apache Hive, Apache Kafka
Database	MongoDB, SQL, MySQL
Machine Learning Tools	TensorFlow, Keras, PyTorch, XGBoost, Naive Bayes, Numpy, Pandas
Workflow Automation	Apache Airflow
Platforms	Windows, Linux
Cloud Services	Azure, DataBricks
Version Control & management	Git, GitLab, Github, JIRA, Agile Scrum, SDLC
Others	Docker, Jenkins, CI/CD, REST APIs, Django, JSON, XML, Microsoft Office

WORK EXPERIENCE

Machine Learning Engineer Intern Jan 2024 - April 2024
BlackBerry, Canada

- Engineered and deployed machine learning models, specializing in **DNN**, **XGBoost**, **RNN (LSTM)** and **LLM**, to analyze and mitigate **cybersecurity threats** within large datasets, enhancing BlackBerry's security infrastructure.
- Utilized **Principal Component Analysis** and **SMOTE** for advanced feature engineering, optimizing representation and mitigating bias. Improved machine learning model performance by 8%.
- Implemented **Local Outlier Factor** and **Isolation Forest** algorithms to detect outliers, enhancing data integrity.
- Leveraged **PySpark** for scalable and efficient data transformations as well as **Integrated streaming data** seamlessly, demonstrating expertise in distributed computing for handling large-scale datasets.
- Collaborated seamlessly with **cross-functional** teams by actively utilizing **Jira** and **Confluence**, ensuring effective project management and streamlined communication.
- Applied **Python** and **SQL Server** for robust data analysis and algorithm implementation, ensuring the identification and prevention of security vulnerabilities, showcasing a commitment to data-driven security measures.

Data Processing Specialist April 2022 - Dec 2022
NeilsonIQ, India

- Developed automated tools using Python, utilizing libraries such as **NumPy**, **Pandas**, and **OpenCV**.
- Spearheaded **database checks** and processing for high-profile clients (PNG, Coca-Cola, Heineken, SPG, Pepsi).
- Designed and accomplished efficient and **scalable automation** solutions to improve database management and processing workflows.
- Collaborated with cross-functional teams to understand client requirements and translated it into automated solutions, ensuring a **15% increase** in timely and **high-quality deliverables**.
- Orchestrated rigorous testing procedures using industry-standard testing tools like **Selenium** and **PyTest** to validate the functionality and reliability of automated tools, ensuring they meet the highest quality standards.
- Enhanced **data quality** by creating a rigorous testing framework that includes **unit and integration tests**. This ensures a reliable and error-free data **processing pipeline**.
- Additionally, incorporated **Power BI** for enhanced **data visualization** and **database insights** presentation.

Big Data Engineer

Jul 2021 - May 2022

Reliance Industries LTD, India

- Performed **Spark jobs** on petabytes of incoming data with **Azure Databricks**, harnessing the parallel processing and fault-tolerance of Spark's distributed computing framework to meticulously analyze large-scale datasets across a spectrum of use cases.
- Demonstrated proficiency in designing and implementing prototypes for various **data architecture solutions**, gaining a deep understanding of different architectural patterns and its applicability to specific business requirements.
- Developed and maintained data tables in **Hive**, enabling efficient data storage and retrieval for analytical purposes, and ensuring optimal query performance.
- Designed the data pipeline architecture using **Apache Airflow** in team of 5 for a new product called JIO Meet that scaled from 0 to **350,000** daily active users.
- Reduced data processing time by 12% by optimizing **ETL workflows** using **Apache Spark** and **Python** and implemented parallel processing techniques using **Hadoop** and **HDFS**.
- Utilized **Azure Cosmos DB**, **Azure SQL Storage**, and **Azure Data Lake** to efficiently manage and store large datasets, resulting in a 20% reduction in storage costs.
- Worked closely with data scientists and analysts to understand data requirements and providing necessary datasets and transformations using services like **Azure Machine Learning** and **Azure Data Factory** for their analytical and modeling tasks.

PROJECTS

Grocery Website Analysis

May 2023 - Jun 2023

University of Windsor, ON

- Implemented a Grocery Website Analysis prototype using **Selenium** for **web scraping** and parsing **HTML** content using **JAVA**, also implemented an inverted indexing approach to validate and extract relevant text from various shop websites in Canada in order to utilize extracted data to provide users with the best deals available.
- Showcased strong problem-solving skills and a deep understanding of e-commerce dynamics.

Comparative Study of Trainable Network in Transfer Learning

Jan 2023 - Apr 2023

University of Windsor, ON

- Developed a project with the goal of optimizing dog breed identification using advanced **transfer learning models** like **RNNs** and **CNNs**.
- Leveraged cutting-edge technologies to modify pre-trained models using **mlflows** through layered adjustments, aiming to boost **efficiency** in **classifying** labeled dog images.
- Advanced **computer vision** applications by refining model training processes for improved performance in dog breed identification.

EDUCATION

Master of Applied Computing

Jan 2023 – Apr 2024

University of Windsor, Canada

Bachelors of Technology in Computer Science and Engineering

Jul 2018 – Mar 2022

Parul University, India

CERTIFICATES

Microsoft Azure Data Fundamentals certification (DP-900)
Software Design: From Requirements to Release
Using Databases with Python

Microsoft
LinkedIn
University of Michigan

Oct 2023
Apr 2023
June 2020