

Himanshu Rao

Linkedin| Github| Portfolio
+1 (437) 988-9899 | himanshurao8999@gmail.com

EDUCATION

UNIVERSITY OF WINDSOR

MASTER OF APPLIED COMPUTING
April 2024 | Windsor, ON
Cum. GPA: 7.9

GALGOTIAS UNIVERSITY

BACHELOR OF TECHNOLOGY IN
COMPUTER SCIENCE
May. 2021 | Greater Noida, India
Cum. GPA: 7.8

COURSEWORK

GRADUATE

Introduction To Data Science
Data Analytics and Visualization
Advanced System Programming
Machine Learning
Information Theory
Internet of Things

UNDERGRADUATE

Python for Data Science and AI
Object Oriented Programming in C++
Data Structures And Algorithms
Programming In Java
Data Science And Big Data Analytics
Data Warehousing And Data Mining
Software Design and Architecture
Database Management Systems
Advanced-Data Analytics
Operating Systems And Linux
Principles Of Cloud Computing
Probability And Queuing Theory
Discrete Mathematics
Computer Networks

SKILLS

TECHNICAL

• Python • Java • C++ • SQL Server •
PySpark • SparkSQL • MongoDB •
Hadoop • Hive • AWS • Snowflake •
HTML • CSS • JavaScript • ReactJS •
NodeJS • DevOps • Jenkins • SDLC • ETL
LIBRARIES/Frameworks
• DJANGO • Spring Boot • Kafka •
OpenShift • Sci-Kit Learn • Matplotlib •
Pandas • Numpy

TOOLS/APPLICATIONS

• Visual Studio • Trino • Docker • JIRA •
Tableau • GitHub • Dataiku • IDL • JCL •
Linux • WebSphere • Excel • Outlook

EXPERIENCE

RBC INSURANCE | CO-OP, FINANCIAL DATA ANALYST

September 2023 – April 2024 | Mississauga, ON

- Designed, developed, and supported batch and real-time data pipelines, optimizing data flow from source data to application database via Real-Time/Batch functions.
- Led the implementation of data processes for Data Warehouse and production systems, ensuring adherence to best practices, regulations, security, and risk standards.
- Created automated python and scala scripts that streamline the month-end survey consolidation process, reducing processing time by 30%.

PUMPUMP PVT. LTD. | DATA ENGINEER

January 2021 – August 2022 | New Delhi, India

- Designed and implemented ETL workflows to extract, transform, and load data from various sources into a Hadoop-based data lake.
- Automated data pipelines using scheduling tools to streamline data processing and improve efficiency.
- Developed scripts in shell scripting, Python, PowerShell, Unix Shell and other scripting languages for automation and data processing.
- Implemented information security and encryption measures to protect sensitive information within the data lake environment.

CIGNEX DATAMATICS | SOFTWARE DEVELOPER INTERN

July 2020 – December 2020 | New Delhi, India

- Utilized Java and Java Frameworks like Spring (MVC, Security, JPA, Boot) to develop robust and scalable applications.
- Designed and implemented a Kafka-based data streaming pipeline to data ingestion process and analyze real-time data streams.

RELEVANT PROJECTS

GOODREADS ETL PIPELINE | DATA ENGINEERING PROJECT

January 2023 – March 2023 | Windsor, ON

- Data is captured in real-time from the Goodreads API using the Goodreads Python wrapper.
- The data collected from the Goodreads API is stored on a local disk and is timely moved to the Landing Bucket on AWS S3. ETL jobs are written in spark and scheduled in airflow to run every 10 minutes.

WEB SCRAPING DATASET FOR SOCCER PLAYERS

SEGMENTATION | DATA SCIENCE AND MACHINE LEARNING PROJECT

September 2022 – November 2022 | Windsor, ON

- This machine learning model achieves a prediction accuracy of 85% for player substitution using advanced feature selection techniques.
- Scrapped a dataset that reflects the most recent player's ratings.
- Created a dashboard that completes a quick analysis of different player's abilities based on various features.

RECOGNITIONS

2023 Completed Machine Learning by Andrew NG on Coursera.
2023 Completed Introduction to Data Science in Python on LinkedIn
2023 Completed Hackerrank Certified SQL Intermediate Level
2022 Completed Introduction to Hadoop and Hive on LinkedIn