

# PUSHPAKRAJ SENTHILKUMARAN

25 the Grantstown park ◇ Waterford, Ireland X91K3H7

(+353) · 892101471 ◇ pushpaksum@gmail.com

## EDUCATION

---

### University Of Limerick , Limerick

August 2025

Master of Science

Major in Artificial intelligence and Machine Learning

Relevant Coursework: Data Analysis; Artificial Intelligence, Deep Learning, Machine Learning, Data Pipeline

Overall GPA: 5.678

### St. Joseph's College of Engineering, Chennai TN,India

Sept 2018 - Jun 2022

Bachelor of Engineering

Major in Artificial intelligence and Machine Learning

Electronics and Communication Engineering

Overall GPA: 8.10

## EXPERIENCE

---

### Cognizant Technologies Solutions

Jul 2022 - August 2024

Big Data Analyst

Chennai TN, India

- Built an end-to-end data pipeline using Kafka, Apache Spark, PostgreSQL, and Grafana
- Ingested raw data and streamed it via Kafka to enable real-time processing
- Processed and transformed data using Apache Spark for efficient batch and streaming analytics
- Stored the cleaned and structured data in Cassandra for downstream access
- Created dynamic dashboards in Grafana to visualize key metrics and insights in real-time

### Dockerized Data Ingestion & Orchestration Pipeline — Python, Airflow, AWS S3, PostgreSQL for development and testing

- Designed and deployed a production-ready, containerized data ingestion pipeline using Docker Compose, integrating Airflow for workflow orchestration and PostgreSQL as the target database.
- Built a modular, multi-container setup including:
  - Airflow DAGs to automate data pulls from public parquet URLs, transform to CSV, and push to AWS S3.
  - A Python ingestion container that retrieves data from S3 and loads it into Postgres.
  - pgAdmin UI for SQL exploration and schema validation.
- Configured custom Docker networks to ensure seamless container communication and maintain environment isolation.
- Leveraged XComs for inter-task communication and enabled cross-DAG data sharing to mimic real-world pipeline dependencies
- Demonstrated proficiency in orchestrating scalable data workflows with built-in retries, logging, and AWS credentials integration using .env and boto3.

### Cognizant Technology Solutions

Jan 2022 - June 2024

Cloud Computing Intern

Chennai TN, India

- Created Pipelines: Developed and configured pipelines to connect APIs to a data lake and Storage containers, ensuring efficient data flow and transformation.
- Implemented Views and External Tables in Synapse Analytics: Designed and created views and external tables in Azure Synapse Analytics, enabling optimized querying and integration of data from various sources.

## UNIVERSITY PROJECTS

---

### Streaming Data to Aws Sagemaker:

- Used Amazon Kinesis to stream real-time customer review data for sentiment analysis.
- Deployed AWS Lambda to process and store transformed JSON files in an S3 bucket.
- Integrated Amazon SageMaker for real-time inference using a deployed Hugging face model MarieAngeA13/Sentiment-Analysis-BERT on incoming reviews.
- Stored the predicted sentiment results in DynamoDB via a triggered Lambda function.
- Configured AWS Glue Crawlers to catalog the DynamoDB table into AWS Glue Data Catalog.
- Utilized Amazon Athena to query the cataloged data efficiently.
- Connected Athena to Microsoft Power BI using the ODBC driver for low-code data visualization and reporting.

### Multiclass image classification using CNN:

- Created Pipelines: Developed and configured pipelines to connect APIs to a data lake and Storage containers, ensuring efficient data flow and transformation.
- Implemented Views and External Tables in Synapse Analytics: Designed and created views and external tables in Azure Synapse Analytics, enabling optimized querying and integration of data from various sources.

## TECHNICAL STRENGTHS

---

<b>Computer Languages</b>	Python, SQL, JavaScript, Bash
<b>Protocols &amp; APIs</b>	REST, JSON, XML, HTTP, AWS SDK (Boto3)
<b>Databases</b>	MySQL, PostgreSQL, DynamoDB, Amazon RDS
<b>Big Data</b>	PySpark,Airflow,AWS Glue, Athena, Kinesis
<b>Machine Learning</b>	AWS SageMaker, scikit-learn, TensorFlow, Keras
<b>Cloud Platforms</b>	AWS (Lambda, S3, EC2, IAM, CloudWatch, API Gateway), Azure (ADLS, Data Factory,Synapse Analytics)
<b>Tools</b>	Git, Docker, Streamlit, Jupyter, AWS CLI, Azure CLI, Postman, Power BI