

Khushi Agarwal

✉ khushi.agarwal21@vit.edu ☎ 9468799699 🌐 khushii-agarwal 🔄 KhushiiAgarwal

Skills

Programming & Scripting Languages: Python, R, PySpark, Java, HTML & CSS, PHP, C, Embedded C, Bash

Systems: Operating System (Windows & Linux), Copilot Studio, Azure AI, Cloud (AWS and GCP), GitHub, GitLab, Minio

Big Data Tools: PySpark, Nifi, Hadoop (HDFS), Airflow, Tableau, Google Looker, Power BI, LLM, Docker, Kubernetes

Database Management: MySQL, PostgreSQL, Oracle, Trino, PrestoSQL, Hive and Iceberg Tables, Google Firebase

Professional Experience

Big Data Intern, Seagate Technology Jul 2024 – Present

Project: Migrate Legacy Hadoop File System to Kubernetes-based Minio system

- Configured and managed **Kubernetes clusters**, optimized **server provisioning**, and handled **pod orchestration**.
- Developed a custom **Data Validation Framework** to ensure data integrity and system reliability after migration.
- Automated data orchestration with **50+ NiFi pipelines** and **50 ETL pipelines through Airflow**, improving migration efficiency.
- Migrated **200+ Hive tables to Iceberg** for enhanced performance, storage optimization, and ACID transaction support.

Project: Maintain and Upscale Enterprise Hadoop Cluster

- Resolved a **5-year-old critical issue**, and enhanced workflows by optimising and deploying **30+ ETL pipelines**.
- Deployed an **ETL Pipeline (Airflow)** using **PySpark** to process **Kafka feeds (JSON → CSV)**, enabling multiple data pipelines.
- Developed **PySpark** and **shell scripts** for seamless CRUD operations across **HDFS, AWS S3, and Minio**.

Project: Custom GenAI Chatbot and BI

- Built the **RAG pipeline, Prompt engineering** and utilised **OpenAI API key** for developing **GenAI chatbot**.
- Generate 4 interactive dashboards and reports using **Tableau** and **Power BI**.
- Utilised **Grafana** and **Prometheus** for proactive resource monitoring, ensuring system health and performance.
- Spearheaded **Young Professionals ERG (India)** and managed events for all Indian sites with 200+ employees.

Leadership and Campus Activities

ExeCom, IEEE SB VIT Pune

- DSA Instructor (Arrays & Stacks) to 300+ students and contributed to AI-ML research using Neural Network, K-NN & Regression, enhancing breast cancer prediction accuracy to 90%+.

Class Representative, FY 2021-2022

- Acted as a liaison between 60 students and faculty, ensuring seamless communication and coordination.


Education

B.Tech in Computer Engineering, Savitribai Phule Pune University (VIT Pune) May 2025 | CGPA: 9.21

12th, CBSE - Central Board of Secondary Education Jun 2021 | Percentage: 95.4%

10th, CBSE - Central Board of Secondary Education Jun 2019 | Percentage: 96.4%

Academic Projects

Federated Learning 

- Built **Federated Learning model** using **Regression (99.2% accuracy)**, ensuring data security & robust model performance.

Pharmacy Inventory Management


- Developed a **Java-based GUI system** to manage medicinal inventory, optimizing selection based on cost and composition.


Achievements


Top 5 Finalist – TensorFiesta Hackathon (Python-based NLP web app).

Academic Excellence: School topper (CBSE) in 10th & 12th grades.


Publications

Amalgam based Multilingual Text Summariser for Devanagari Languages, IEEE (Scopus Indexed & IEEE Xplore) 

Federated Learning approach to detect Arrhythmia, IEEE (Scopus Indexed & IEEE Xplore) 

AI Based Networking, Software-Hardware Co-design for AI/ML, Blog on Medium 

Certificates

- Generative AI for Developers by Google 
 - Google Data Analyst using Looker 
 - Fundamentals of Deep Learning by Nvidia 
- Data analysis in Power BI by Microsoft 
 - IBM Full Stack Software Developer 
 - IBM Back-End Development 