

# Ashval Vaidya Data Scientist

✉ ashvalvaidya@gmail.com    📞 8806949268    📍 Pune, India    🔗 LinkedIn    🐙 Github

## Profile

Data Scientist with 8.5 years of experience, including 4 years in AI, machine learning, and predictive analytics. Skilled in leveraging large language models (LLMs), text embeddings, and clustering to deliver scalable solutions. Experienced in building knowledge retrieval systems, predictive maintenance models, and generating actionable insights across energy and IT infrastructure domains. Proficient in cloud platforms (AWS, Azure), big data tools (Spark, Hadoop), and SQL. Strong focus on applying data-driven approaches to solve complex business problems through efficient data extraction, cleaning, and analysis.

## Education

**Msc in Big Data Science, IESEG School of Management** Sep 2022 – Dec 2023 | Lille, France  
**Bachelor of Technology in Computer Science,** 2011 – 2015 | Mumbai, India  
*Narsee Monjee Institute of Management Studies*

## Skills

**Languages:** English (Fluent), French (Beginner)  
**Programming:** Python, R, SQL, PySpark, SAS  
**Data Visualization:** Tableau, Power BI, Excel, Streamlit, R Shiny  
**AI, Machine Learning & NLP:** LLMs, NLP, Deep Learning, Text Embeddings (GUSE, BGE-4, Sentence Transformers), Supervised Learning, Forecasting, Clustering, Recommendation Systems, Predictive Modeling, Statistics, scikit-learn, TensorFlow, PyTorch  
**Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, TensorFlow, PyTorch, Keras  
**Big Data & Data Engineering:** PgVector, Apache Hive, Spark, Hadoop  
**Cloud Services:** MS Azure, GCP, AWS  
**Version Control:** Git  
**Soft Skills:** Analytical Mindset, Attention to Detail, Adaptability, Communication, Collaboration, Empathy

## Professional Experience

**AI Engineer, ALIGNED AUTOMATION (Client: DELL Technologies)** Nov 2024 – present | Pune, India

- Built an Agglomerative clustering model using Dell’s IT support data (60K+ cases) to recommend relevant KB articles, improving resolution time by 30%.
- Used GUSE and BGE-4 embeddings to vectorize case data and implement semantic similarity search.
- Stored embeddings in PgVector; applied FlashRank reranking with 91% top-5 accuracy.
- Collaborated with US-based data scientists and Dell experts.
- Extracted, cleaned, and analyzed data from multiple sources to provide actionable support insights.
- Technologies used: Python, Pandas, scikit-learn, GUSE, PgVector, Git, SQL, AWS.

**Data Scientist, FLOCK ENERGY (Startup)** Dec 2023 – Oct 2024 | London, United Kingdom

- Predicted solar adoption from smart meter data of 200,000 households with 93% accuracy using Random Forest in PySpark.
- Developed scalable pipelines in PySpark, Python, and SQL to disaggregate appliance-level power usage (e.g., ACs, water heaters), leading to ~\$500/month savings.
- Built Tableau dashboards for customer segmentation and energy consumption trend analysis.
- Mentored junior data scientists, enhancing documentation, coding standards, and team workflows.
- Tools: PySpark, Python, SQL, Tableau, Git, GCP.

**Data Scientist, CINIONIC, A Barco Company** Jan 2023 – Jun 2023 | Kortrijk, Belgium

- Led a data-driven initiative involving in-depth analysis of 340 Million rows of real-time sensor data of cinema projectors, utilizing PySpark on Databricks. Engineered a high-precision machine learning model with a 75% recall, resulting in a substantial 67% reduction in revenue loss by predicting and preventing projector failures.
- Identified and addressed root causes of projector failures, specifically targeting the top 5 sensors which significantly reduced unscheduled cinema projector downtime, enhancing operational efficiency within the company.
- Regularly interacted with the business team, presenting critical insights through sophisticated PowerPoint presentations and Excel Dashboards. Demonstrated proficiency in data processing using Excel Power Query and Python, with the development of automation scripts streamlining essential processes.
- Tools: PySpark, SQL, Excel, Azure, Hadoop, Git.

**Data Scientist, ZEETRON Pvt Ltd**

Sep 2021 – Aug 2022 | Jaipur, India

- Cleaned and transformed retail sales data, improving model accuracy by 20%.
- Built machine learning models to predict sales and manage inventory using Pandas and Scikit-learn.
- Deployed models on AWS (S3, EC2) for real-time inventory insights.

**Data Analyst, EXPLEO (Client: Trafigura)**

Jul 2019 – Aug 2021 | Pune, India

- Led the Testing initiatives for the Commodity Trading client, executing over 400 SQL scripts ensuring data reliability
- Boosted stakeholder efficiency by 30% by implementing dashboard automation using Tableau and SQL.

**Data Analyst, ACCENTURE (Client: Merck Pharmaceuticals)**

Jan 2016 – Jun 2019 | Pune, India

- Utilized Excel & SQL to drive analysis and automation, demonstrating proficiency in both Agile and Waterfall methodologies.
- Analyzed financial services sales data of 10,000 customers, delivering weekly reports as per the client requirement.

**Certificates**Python PCEP-30-02 [🔗](#)Tableau Desktop Specialist [🔗](#)SAS Base Programming 9.4 [🔗](#)**Academic Project**

- Used Tableau to design dashboards for NYC Flights data, perform calculations and reported flight delays via the ETL process.
- Scrapped Yahoo finance stock data with Python and represented it via Streamlit dashboard based on the user-defined ticker.
- Created a benchmarking prediction model using PySpark to predict customer review ratings to improve the customer experience for an e-commerce company, resulting in an increase in sales by 5%
- Developed a predictive model to identify potential donors for fundraising campaigns increasing the number of donations by 20%.
- Created a marketing DataMart with R and visualized insights via Shiny dashboard using the BWIN Gambling dataset.
- Used NLP to perform advanced sentiment analysis to extract insights from patient's online reviews about drugs and treatment
- Developed a Recipe recommender system that recommends ingredients based on past recipe review ratings.