**ABHISHEK SHAH, PMP**

**Data Scientist**

Portland, OR 97133 • abh2050@gmail.com • + 1 (503) 278-1550 [LinkedIn](http://linkedin.com/in/abhishek175) | [Medium](https://medium.com/@jwbtmf) | [GitHub](https://github.com/abh2050/searchengine) | [Website](https://abh2050.github.io/)

**PROFESSIONAL EXPERIENCE**

**BMW Remote, Or**

**Data Scientist Oct 2024-Present**

* Designed & deployed an XGBoost-based classification model via DataRobot for batch inference to detect paint chip defects in BMW’s paint line. Achieved 92% accuracy and 82% recall, reducing monthly scrap costs by $70K.
* Implemented monitoring for data drift and model drift, ensuring sustained performance and increased production throughput by 5 cars/day, generating $300K/month in addn. revenue through optimized defect reduction.
* Developed a time-series SARIMAX forecasting model to predict vehicle delivery cycle times and deployed it on DataRobot for batch inference. Reduced delivery delays by 30% and optimized inventory management processes. Improved transportation planning, resulting in a 15% boost in revenue realization.
* Developed a new RAG-based legal assistant for contract clause comparison using Claude 3.5 Sonnet on AWS Bedrock. Legal teams (25 attorneys) used it to compare MSAs and vendor contracts. Combined semantic search and prompt engineering to reduce hallucinations. Resulted in 15% faster processing time and saved 10 hours/week per team by automating repetitive review tasks.
* Spearheaded training for BMW’s proprietary Group AI Assistant, leveraging generative AI capabilities like knowledge graph integration, custom app building, and secure connectors. Trained ~150 employees weekly on AI use cases and workflows, enhancing AI literacy, streamlining operations, and driving generative AI adoption.

**INTEL Hillsboro, Or**

**Data Scientist/Engineer Nov 2019-Oct 2024**

* Built a new RAG pipeline using OpenAI API and FAISS to analyze five years of historical Statements of Work (SOWs) for 45 engineers. Used chunking, embedding, and semantic search to retrieve and summarize relevant scopes. Empowered engineers to generate scopes in minutes instead of hours, saving 1,800+ hours/month and enabling $100K/month in soft cost savings.
* Developed and deployed a machine learning solution using Random Forest for anomaly detection and PCA for dimensionality reduction to identify defects in the FOUP transfer process; optimized the model with OpenVINO and deployed on Intel DevCloud, reducing scrap wafers and saving the company $250K annually.
* Collaborated with Audi to implement an AI-driven quality control system in their Neckarsulm factory, processing data from 900 welding robots to perform real-time, inline inspections of over 5 million daily welds. This scalable solution reduced labor costs by 30% to 50% moving Audi closer to Industry 4.0 ([Project Link](https://www.intel.com/content/www/us/en/customer-spotlight/stories/audi-automated-factory.html))
* Developed a Power BI dashboard using Azure Data Factory (ADF) analytics to monitor 4,000+ pipelines, ensuring real-time tracking of metadata and performance metrics. Visualized pipeline metadata, enabling early failure detection and saving the core analytics team 1 hour daily in troubleshooting data flow issues.

**WESTLAKE CHEMICALS LONGVIEW, WA**

**Staff Data Analytics Process Engineer September 2017-November 2019**

**TATE & LYLE DULUTH, MN**

**Data Analytics Process Engineer May 2012 – August 2017**

**EDUCATION**

**University of Michigan Dearborn, MI**

***Master of Artificial Intelligence and Machine learning; Concentration: NLP* 2022-2025**

**Washington state university Pullman, WA**

***Master’s in Engineering and Technology Management* 2018-2020**

**University of MINNESOTA Duluth, MN**

***Bachelor of Science in Chemical Engineering* 2007-2012**

**Tech Stack**

Scikit-learn, TensorFlow, DataRobot, AWS, Azure, GCP, Claude, OpenAI API, PCA, OpenVINO, ADF, Power BI, Python, MongoDB, T-SQL, Airflow, PySpark, PostgreSQL, Docker, Kubernetes, Git, Qlik