ARUNBH YASHASWI

College Park, MD | +1 240 438 1879 | arunbhy@umd.edu | linkedin.com/in/arunbh-yashaswi/ | arunbhy.github.io

EDUCATION

University of Maryland — College Park, MD, USA
Master of Science in Data Science | GPA 4.0/4.0 — *Aug. 2024 – May 2026*
Vellore Institute of Technology — Chennai, Tamil Nadu, India
Bachelor of Technology in Computer Science and Engineering | GPA 3.7/4.0 — *Jul. 2017 – Jun. 2021*

EXPERIENCE

Data Scientist — Mar. 2023 – Jun. 2024
UnitedHealth Group — *Noida, Uttar Pradesh, India*

- Developed and deployed **data-driven automation pipelines** for **unstructured document extraction** within TESS, later integrated into UHAIS, improving workflow automation.
- Implemented **advanced data mining techniques** to enhance **accuracy and efficiency**, leveraging **big data processing** for scalable analytics.
- Optimized **entity extraction models**, increasing accuracy by **18%** and enhancing **computer vision workflows** with **Azure PaaS**, reducing manual efforts by **55%** through modularized services.
- Increased **document matching accuracy to 84%** with **14% faster processing**, leveraging **transformer based models (MPNet)**.
- Processed **45,000+ documents monthly** using TESS-powered ETL pipelines, improving decision-making.
- Accelerated large-scale data processing, reducing computational time by **40%** with **TensorFlow and OpenCV**.
- Led **seamless database migration** across platforms by automating query conversion with **LLMs**.
- Worked with **cross-functional teams**, managing **feature prioritization** while following **Agile methodologies**.

Associate Data Scientist — Jul. 2021 – Mar. 2023
UnitedHealth Group — *Gurgaon, Haryana, India*

- Developed **production-ready machine learning solutions** for **document parsing and its processing**.
- Optimized **OCR-based data extraction pipelines**, reducing processing time by **57%** through **parallelized workflows and concurrency**.
- Enhanced **scalability of big data systems**, improving **system capacity by 380%** via **load balancing and distributed computing**.
- Improved **document structure analysis** with **87% accuracy**, leveraging **DocLayNet and deep learning**.
- Transformed **prototypes into scalable AI models**, achieving **96% accuracy** and **42% faster processing**.
- Designed real-time monitoring dashboards for **drift analysis**, tracking 8 production models across multiple teams.

PERSONAL PROJECTS

LiveScreen Translation | *Python, OpenCV, Tesseract, YOLOv8, NLP* — Jun. 2022 – Present
- Developed a **real-time AI-driven OCR pipeline** for text extraction, language translation, and seamless reintegration into images. For a seamless translation of comic books.
- Optimized **computer vision models** to detect text regions, achieving **82% accuracy** in multilingual translation.

Automated Ticket Creation for email | *Python, NLP, Spark, Databricks, SQL* — Jan. 2019 – Dec. 2019
- Built an AI-driven NLP pipeline for **automating ticket classification and response generation**, capable of handling 20+ requests in single instance per second. With the feature of reverting back to collect missing details.
- Optimized **ETL workflows in Databricks**, reducing **large-scale data processing time by 20%**.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, R, Spark, C++, MATLAB
- **Machine Learning and AI:** Supervised and Unsupervised Learning, Predictive Modeling, Deep Learning, Neural Networks, NLP, Anomaly Detection, Time Series Forecasting, Clustering, Large Language Models (LLMs)
- **Big Data and Data Engineering:** ETL Pipelines, Data Cleaning, Feature Engineering, Data Warehousing, Distributed Computing, Cluster Optimization, Databricks, Apache Spark, Hadoop, Hive
- **Cloud and MLOps:** AWS (S3, SageMaker, Glue, Lambda, Redshift), Microsoft Azure (Synapse, Cognitive Services), Kubernetes, Docker, Delta Lake, CI/CD Pipelines, Model Deployment, SAS
- **Tools and Frameworks:** PyTorch, TensorFlow, Scikit-learn, XGBoost, LightGBM, MLflow, OpenCV, Pandas, NumPy, Matplotlib, Seaborn, Jupyter, Tableau, Power BI

HONORS AND ACHIEVEMENTS

- Patented, engineered cost-effective document extraction boosts scalability for 30+ teams (UHG, Jun. 2023)
- Earned the Performance Bravo for leading scalable AI solutions, automating key operations (UHG, Jun. 2022)
- Secured First Prize at SIH 2019 for developing a Ticket Creation Solution, leading a team of six (Mar. 2019)
- Organized E-sports during Under-graduate for games like CS:GO, R6 and Fortnite. (Feb. 2019)