

Amol Khade

CONTACT

- khadeal@mail.uc.edu
- (513) 652-2242
- [LinkedIn](https://www.linkedin.com/in/khadeamol/) (linkedin.com/in/khadeamol)
- [GitHub](https://github.com/khadeamol/) (github.com/khadeamol)

EDUCATION

- **MS Information Systems**
University of Cincinnati, USA (2018)
- **BE Computer Science**
University of Pune, India (2012)

TECHNICAL SKILLS

- Proven record of working with Python for **API and SDK Development**, Data Engineering and Software Engineering (Object Oriented architecture), Machine Learning and Data Science
- Excellent SQL skills
- **Python libraries:** Boto3, Vaex, PyTorch, PyArrow, PySpark, DuckDB, Pandas, FastAPI, Scikit-learn, Pandas, NumPy, Keras, TensorFlow, SQLAlchemy, Dask
- CNNs for image recognition
- Experiment tracking with MLFlow
- dbt, Airflow, Docker, Kubernetes, Jenkins for CI/CD, container management, deployment, workflow orchestration and monitoring
- Model deployment via Kinesis, Lambda, MLFlow
- **Data/BI Platforms:** Dremio, Alation, Looker,
- **Advanced data visualization, dashboards and reporting:** Tableau, Birst, Seaborn, Matplotlib
- Worked on AWS ecosystem: EC2, S3, Athena, Redshift
- Time series analysis, statistical methods and techniques

CERTIFICATIONS

- **Stanford Supervised Machine Learning** by Andrew Ng (July 2022)
- [IBM Certified Data Scientist](#)
- [Data Science with Python: Codecademy](#)
- [Machine Learning in Python: Codecademy](#)
- [Data Science, Machine Learning in R Udemy](#)
- [Advanced Google Analytics](#)

PROFILE SUMMARY:

Through my varied experience being a part of organizations of various scales over the past 6 years, I have gained the skills to be an all-round asset to any data team. I have experience working with pure data analytics, data modelling, data engineering as well as software engineering at my last engagement. Alongside these, I have also learned Data Science skills. I am confident of being a valuable member of any team.

DATA SCIENCE & MACHINE LEARNING PROJECTS

End-to-end Machine Learning Pipeline

- [Notebook](#): Libraries used: Scikit-learn, NumPy, Pandas
- Calculate efficiency (MPG) for car – given its other attributes ([dataset](#))
- Used scikit-learn methods: Stratified Shuffle split, Train Test split, One Hot encoding, Simple Imputer, Standard Scaler, Column Transformer
- Implemented pipelines for data transformations, creating custom attributes through custom functions
- Models evaluated: Linear regression, Decision Tree regressor, Random Forest regressor, Support Vector Machine regressor
- Model performance evaluation using Root Mean Squared Error
- Implemented Hyperparameter tuning and K-fold Cross Validation

Housing Market Prediction

- Enables real estate firms to design pricing strategies to maximize revenue ([Notebook](#))
- Libraries used: Scikit-learn, NumPy, Pandas, Matplotlib, Seaborn

E-commerce Customer Segmentation Analysis

- Used K-Means Clustering Frequency, Recency, Monetary Value (FRM)
- Compared model performance based on inertia and number of clusters
- Identified Customer outreach strategies based on cluster metrics

Image Classification using PyTorch (CIFAR-10 dataset)

- Built Convolutional Neural Network Model using PyTorch
- Calculated gradients for backpropagation on Neural Networks
- Utilized Tensors for images in Python Image Library (PIL)

EXPERIENCE

Data Engineer | xFusion Technologies

February 2022 – March 2023 | Sacramento, CA

- Product: Low-code ETL pipeline automation
- Engineered Python APIs and microservices for data processing, analytics, data science features on data in PostgreSQL, S3 and Snowflake
- Incorporated Vaex in the tech stack to dramatically improve processing times by up to 80%, and reduced computing resource demand
- Developed custom Airflow operators
- Created Python API/SDK based functionalities:
 - **AWS S3 Bucket Explorer:**
 - Self-serve feature to manage content in S3 buckets
 - Ability to view/analyze data from within the platform UI
 - **Data Quality component:**
 - Generates statistical data quality metrics on customer data
 - **SQL Query Execution Engine:**
 - Allows user to run queries on data in S3 using DuckDB
 - Creates JSON payload with results and performance metrics

Data Management Analyst - Data Engineering | Avant Financial, LLC

October 2019 – June 2021 | Chicago, IL

Data Asset Migration:

- Ownership of Credit Card Application data migration
- Designed and owned data models and ETL processes for migrating datasets using Airflow, PySpark to eliminate redundancy by up to 50%
- Ownership of ETL processes, collaborating with other engineers, business teams to address issues affecting data integrity, availability
- Delivered **Looker** Looks, Dashboards, Queries for business users to streamline reporting. Saved human labor cost for the organization by 75%

(Continued on next page)

Amol Khade

Data Analyst - Data Science and Strategic Analytics | Gloo, LLC

June 2018 - July 2019 | Boulder, CO

- **Spearheaded identification and exploratory analysis** on new data sources to add value [Gloo's Insights platform](#) and enhancing audience coverage across the country and demographic
- **Created automated auditing scripts in R** for periodic reporting on high volume datasets, capturing longitudinal trends in an evolving data set
- Developed an **interactive Tableau dashboard** for a client to capture sentiment analysis on Twitter content

Data Analyst - Intern | Crossroads Community Church

June 2017 – May 2018 | Cincinnati, OH

- Performed as the **Subject Matter Expert for Tableau** trained the team on using Tableau as their primary BI tool
- **Authored interactive Birst dashboards** and setup a self-service BI reporting mechanism for stakeholders, eliminating analytics bottlenecks team. Explored fund allocation opportunities for the finance team