ARGHYA MUKHERJEE

Tulsa, Oklahoma
m: 918.330.1803
arghyamukherjee1991@gmail.com
GitHub | Scholar | LinkedIn

OBJECTIVE

To advance my career with a leading organization that will utilize my extensive data engineering, data science skills experience focused on econometrics analysis, computer vision and Generative AI to transform the future.

Proven expertise in building and optimizing scalable data pipelines to solve complex business challenges and deliver transformative insights. A detail-oriented data engineer with a strong foundation in data architecture, ETL processes, and cloud-based data solutions, dedicated to leveraging innovative technologies to drive impactful data-driven solutions. Adept at synthesizing large datasets, implementing robust data workflows, and collaborating across teams to ensure alignment with strategic objectives. A motivated and growth-oriented professional committed to driving value and achieving measurable business outcomes through efficient data science practices and cross-functional collaboration.

PROFESSIONAL EXPERIENCE

WORKONWORD
DATA SCIENTIST/ AI ENGINEER MAY 2024 – PRESENT [AWS BEDROCK, SAGEMAKER, S3, EC2, PYTORCH, PYTHON, GITHUB WORKFLOWS]
- Design and implement A/B testing on website and app and implement suggest design changes.
- Worked with stakeholders in designing ETL/ELT pipelines used for building predictive models.
- Developed a proprietary algorithm using feature engineering, deep learning and generative AI to automate the matching of resumes with job descriptions, enhancing recruiter efficiency.
- Managed a team in designing and implementing an AI chatbot for recruiters to interact with resumes in real time, enabling detailed queries and retrieval of candidate insights from large databases.
- Created backend data classification using unsupervised machine learning to streamline the organization of incoming data improving productivity and decreasing cost of operation.
- Developed a GenAI model which provides descriptive job description based on key information resulting in 70% more number of recruiters using our platform to generate job descriptions quickly. and efficiently increasing revenue by 50% since the launch of the model

ANTI-PHISHING WORKING GROUP, REMOTE, OCTOBER 2020 – PRESENT [PYTHON, R, AWS, ELASTIC SEARCH]
SOFTWARE ENGINEER (DevOps)
- Formed Rest APIs for Cryptocurrency Working Group clients from both the academia and industries.
- Garnered, reviewed, and analyzed data and updated a database of approximately 1.2million BTC wallet addresses used for scams and ransomware.
- Tested systems to measure reliability and efficiency ensuring 99.999% uptime for clients.
- Agile software development method implemented resulting in quick turn around and fast feature development using CI/CD pipelines using git, Jenkins and Apache airflow to manage workflow.

ROBBIE AI (ATENTO CAPITAL PORTFOLIO) [KAFKA, AWS, AZURE, GIT, PYTHON, OPEN-CV]
DATA SCIENTIST /DATA ENGINEER MAY 2023 – APRIL 2024 (FULL-TIME)
- Engineered scalable data pipelines using Apache Airflow, integrating cloud storage solutions with Apache Kafka for efficient real-time video and data processing, ensuring seamless data flow between systems.
- Integrated advanced deep learning models into production pipelines, focusing on optimizing the data flow and enhancing model efficiency for computer vision-based body structure analysis.
- Achieved over 23% improvement in model performance through targeted optimizations, focusing on pipeline efficiency and data handling strategies to improve the accuracy of predictions.
- Developed forecasting models for predicting patient falls, using in-house models and deploying them on Nvidia Jetson devices with CUDA support, reducing nursing costs by nearly 30% for clients.

THE UNIVERSITY OF TULSA, TULSA, OK (FULL-TIME) [PYTHON, R, GIT, AWS SAGEMAKER, S3, SNOWFLAKE, SIEM]
DATA SCIENTIST, JANUARY 2018 - PRESENT
- Developed and maintained ETL pipelines to extract, transform, and load data from various blockchain networks, focusing on Bitcoin, Ethereum, and Solana. Automated data ingestion processes using Python, Snowflake, and SQL to ensure real-time data availability.
- Managed and optimized data storage solutions for high frequency blockchain data, ensuring secure and efficient access for research and analysis.
- Conducted research and experiments to build hypothesis and build estimation models using multiple logistic regression, followed by predictive models using Neural networks.
- Used NLP, Generative AI to classify large chunks of texts into different categories efficiently by fine tuning models.
- Analyzed complex datasets using statistical tools like Stata and R and built predictive models to measure the failure and stability of exchanges and cryptocurrencies. Focused on identifying market trends and price movements across blockchain ecosystems.
- Worked closely with cross-functional teams, sharing findings and insights through detailed reports and presentations on market trends and blockchain ecosystem behaviors as well as bridging the gap between economics and cyber security by measuring breaches and building optimal security investment models.

PRICEWATERHOUSECOOPERS, BIRMINGHAM, UK, FEBRUARY 2015 - DECEMBER 2016 (FULL-TIME)
IT RISK ADVISORY ASSOCIATE
- Participated actively in the delivery of IT risk advisory services comprised of risk assessments, security reviews, compliance assessments, and IT governance assessments.
- Studied and evaluated client IT environments to ascertain risks and control gaps and provided recommendations for remediation solutions.
- Test systems to validate compliance according to local regulations.
- Cross-collaborated with multiple teams to provide IT risk advisory services, prepared reports and presentations, and maintained knowledge of current regulatory and industry requirements.

EDUCATION AND CREDENTIALS

PHD IN COMPUTER SCIENCE, MAY 2024
> *The University of Tulsa - Tandy School of Computer Science, Tulsa, OK, GPA 3.67*
> *Dissertation: Measuring cyber security shocks and scams in cryptocurrency eco-system*

POST GRADUATE DEGREE IN COMPUTER SCIENCE, NETWORK, AND DISTRIBUTED SYSTEMS, 2014
> *University of St. Andrews - Department of Computer Science St. Andrews, Fife, UK, GPA 3.33*
> *Graduated with GPA 3.3 Equivalent*

UNDERGRADUATE DEGREE IN COMPUTER APPLICATIONS, 2012
> *West Bengal University of Technology, Kolkata, India, GPA 3.33*

TECHNICAL EXPERTISE

LANGUAGES: R, Python, Java, C/C++, MATLAB, Latex, JavaScript, ReactJS, Solidity, SQL, NoSQL
TOOLS: Power BI, Tableau, Linux, Azure, AWS, Kafka, GCP, Web3, Docker, Selenium, PyTorch, Tensorflow, Keras, PyCaret, Figma, Jira, Notion, Redshift, Athena, Sagemaker, Django, Flask, Tornado, SAS, Stata, Jenkins, CI/CD, Fast API, Transformers, Hugging Face, Jenkins, Git, PySpark, NumPy, Pandas, Lambda, DynamoDB, AWS bedrock, Lambda
SKILLS: Statistics, Survival Analysis, Time Series Analysis, ML, NLP, Quantitative Research, GraphQl, Rest API