NISCHAY THAPA

DATA SCIENCE & ANALYTICS

CONTACT

- nischaybikramthapa14@gmail.com
- (040) 374-9059
- Melbourne,VIC
- linkedin.com/in/nischaythapa
- github.com/nischaybikramthapa

PROFILE

Data Scientist and Consultant for Public Sectors at The Data Foundry.

EDUCATION

2021

RMIT UNIVERSITY

Master of Data Science

2018

TRIBHUVAN UNIVERSITY [NEPAL]

Bachelor of Information Management

MANAGEMENT SKILLS

- Agile Development
- Critical Thinking
- Communication & Storytelling
- JIRA, Miro, Git, Lucid App

CERTIFICATIONS

- AWS Certified Cloud Practitioner
- AWS Solution Architect Associate

TECHNICAL SKILLS

- **AWS** – Lambda, Kinesis, Glue, Databrew, Step functions, DynamoDB, Redshift, Sagemaker, CodePipeline, Athena, Quicksight, Batch, ECR, API Gateway, Cloudformation
- **Machine Learning** – Supervised, Unsupervised, Self-supervised
- **Data Engineering** – Accelerated Data Lakes, ETL / ELT Pipelines, Data Quality Pipelines
- **Natural Language Processing**- Text Mining, NER, Topic Modeling, Q/A
- **Data Visualisation** – Plotly, Tableau, PowerBI, Quicksight
- **MLOps:** MLflow, Airflow, Docker
- **Tools** - Python, R, SQL, JAVA, Scala, Spark

EXPERIENCE

2021 - PRESENT

Data Scientist | The Data Foundry

Current Projects

- Designing an **ingestion pattern** to fetch data from various social media into the data lake to power **marketing analytics** for RMIT University.
- Developing a modular **MLOps framework** to support **real-time inference** and batch workloads, monitor **concept** and **data drifts**, cloud watch alarms to **re-train ML pipelines** and **feature store** using Sagemaker, CodePipeline, API Gateway, Lambda and CDK.

Past Projects

- Delivered enterprise **data lake solution**, player performance reporting tool for Golf Australia, improving their analytical capabilities by **80%.**
- Developed an **ETL workflow** to automate the transformation of deeply **nested JSON documents** for cataloging in AWS Glue, significantly reducing the effort to analyse the data by the end user.
- Re-engineered and automated legacy ingestion and consumption pattern by using **AWS Lambda with a customised R runtime** thereby achieving considerable cost saving and improved performance.
- Used **CTEs** and advanced **analytical function** to chain multiple data sources and identify high potential athletes, improving Athena query performance by **200 seconds** while scanning **more than 500 GB** data.
- Developed a **curation engine** to run multiple Athena queries on a schedule using Eventbridge while storing metadata in DynamoDB, powering more instantaneous reporting capabilities.

2021–2021

Data Scientist | PHM Technology

- Developed an **active learning** framework to reduce **labelling cost** in regression and classification problems for **safety** and **mission critical** systems.
- Implemented **custom query strategies** for different learning scenarios, improving model performance by **17%** than random selection using only **40% of training data**.
- Built an interactive **CT/CD pipeline** as a prototype using **streamlit** and **mlflow** to acquire labels from an oracle, conduct multiple experiments and serve trained model into staging and production.

PUBLICATION

Hospital Readmission Prediction Using Clinical Admission Notes | *Association for Computing Machinery*

Proposed a novel approach to identify **early readmission risk** of patients using clinical **admission text data** from electronic health records.