

NISCHAY THAPA

LEAD DATA SCIENTIST

CONTACT

- ✉ nischaybikramthapa14@gmail.com
- ☎ (403) 749-059
- 🌐 Melbourne, Australia
- 🌐 linkedin.com/in/nischaythapa
- 🌐 github.com/nischaybikramthapa

PROFILE

Data Scientist with over 5 years of experience in data engineering, data science and advanced analytics on various projects for higher education, high-performance sports, state government, healthcare, and telecommunications.

EDUCATION

2021
RMIT UNIVERSITY
Master of Data Science

2018
TRIUVAN UNIVERSITY [NEPAL]
Bachelor of Information Management

MANAGEMENT SKILLS

- Agile Development
- Critical Thinking
- Leadership
- Project and Team Management
- JIRA, Git, Lucid App

CERTIFICATIONS

- AWS Machine Learning Specialty
- AWS Data Analytics Specialty
- AWS Solutions Architect Associate
- AWS Certified Cloud Practitioner
- Databricks Lakehouse Fundamentals

EXPERIENCE

2021 - PRESENT

Lead Data Scientist | The Data Foundry

Leadership

- Leading a team of 5 data scientists delivering machine learning solutions for a diverse range of industries (healthcare, education, government, and high-performance sports).
- Produced **architecture** and statement of work documents for project tenders and contracts totaling approximately \$2 million.
- **Driving strategy and visions** for machine learning by translating business use cases into innovative solutions for customers.
- Collaborating with data engineering and data analytics teams to deliver end to end machine learning products with a focus on **data quality, performance, scalability, and automation**.
- Engaging with customers to gather valuable feedback and developing **machine learning accelerators** to enhance delivery speed and promote reusability.

Past Projects

- Lead the data science team to identify golfers that are at **risk of churn** using XGBoost trained on AWS, helping Golf Australia increase their **retention by 7%**.
- Developed ELT pipelines for RMIT University's Data and Analytics Platform using S3, **Snowflake and dbt** to recover **\$2 million** in lost revenue from the past year.
- Developed an **MLOps framework** using AWS Sagemaker for RMIT University to continuously train and predict Early Warning Signs of student attrition with batch scoring, monitoring, re-training and explainability features.
- Built **streaming analytics engine** using Kinesis Data Analytics with Flink to enable real-time decision making for New Zealand's team at the Bermuda Sail Grand Prix, helping them secure **sixth position** at the race.
- Architected and developed a **data quality pipeline** on AWS using Great Expectations and Deequ, significantly improving data consistency, completeness, and accuracy.
- Delivered **enterprise grade data lakehouse** and player performance reporting tool for Golf Australia, reducing reporting time from six months to 2 days.
- Developed an ETL workflow to **automate the transformation** of deeply nested JSON documents for cataloging in AWS Glue, enabling self-service analysis by the end user.

TECHNICAL SKILLS

- **Machine Learning** – Logistic Regression, SVM, Nearest Neighbors, Ensemble methods, Clustering, Collaborative Filtering, Matrix Factorization
- **Deep Learning** – CNN, RNN, Transformers, GAN
- **NLP** – Text mining, Sentiment and Emotion Analysis, Question/ Answering, Name-Entity Recognition, Text summarization, Topic Modeling
- **Statistics** – A/B Testing, Hypothesis testing, Distributions
- **Data Visualization** – PowerBI, Tableau, Quicksight, Plotly, ggplot
- **Data Engineering** – ETL, ELT, Streaming pipeline, data quality, data governance, cataloging, data lake, data lakehouse, data modeling
- **AWS** – Lambda, ECS, ECR, Batch, Fargate, DynamoDB, Redshift, Sagemaker, Glue, Stepfunctions, Kinesis, CDK, Codebuild, Eventbridge, Textract
- **Software Engineering** – Object Oriented Design, Unit Testing
- **Tools** – Python, R, SQL, JAVA, Scala, Spark, Flink, Kafka, Scikit-learn, Great Expectations, dbt, Pytorch, TensorFlow, Docker, Airflow, Prefect

AWARDS

- **3rd Position in IEEE ISDAIOT Symposium** – Proposed a novel framework for public health officials to identify potential groups for vaccination prioritization in the U.S using Hierarchical Clustering and Susceptible, Infectious and Recovery (SIR) model.
- **Digital Innovator** – Developed a Minimal Loveable Product for Museum of London to communicate forecasted electric consumption and potential faults in HVAC Systems.

EXPERIENCE

2021 - 2021

Data Scientist | PHM Technology

- Developed an active learning framework to **reduce data labeling costs** in regression and classification problems for safety and mission-critical systems.
- Implemented customer querying strategies for different learning scenarios, **improving model performance by 7%** than random selection using only 40% of training data.
- Built an interactive **Continuous Training / Continuous Delivery (CT/CD)** pipeline as PoV using streamlit, FastAPI, and MLFlow to acquire data labels from oracle, conduct multiple experiments and served trained models into production.

2017 - 2020

Data Analyst | Worldlink Communications

- Developed a **dashboard using PowerBI** highlighting KPI such as prompt time, retention rate, and customer satisfaction score from more than 400 thousand telecommunication customers to support data-driven decisions.
- Segmented customers using various clustering techniques based on their reported issues, geocode and other attributes to accelerate field visit for maintenance, **improving retention rate by 7%** over six months.
- Analyzed text-based sentiments from user complaints and ranked frustrated customers for follow ups, **improving customer satisfaction score by 14%** over a year.

PUBLICATION

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

- Proposed a novel approach to identify **hospital readmission risk** of patients and support healthcare practitioners for early medical intervention and effective discharge planning.
- Trained **Clinical BERT** on patient's data from MIMIC-III to create vector representation of text and predicted hospital readmission rate with **AUC score of 0.72**, 4% better than the baseline model.
- Provided **interpretable results** using BertViz, SHAP and LIME to help clinicians understand model's predictions.