https://andwct.github.io/
https://www.linkedin.com/in/andwct

ANDERSON WU

0912675082
andwctneuro@gmail.com

INDUSTRY EXPERIENCE

Machine Learning Engineer **TSMC AI4BI – AI Technical Project** **Jan, 2023 – now**

Wafer Start Analysis (Team lead of 4 engineers and 4 interns)

- Data Analytics: Spearheaded a novel cycle time forecast analytics framework for Fab 18A (TSMC leading Node Fab), achieving a 2.8% MAPE, outperforming the benchmark method's 4.5%, and achieving a correlation coefficient of 0.65 compared to 0 with the benchmark.
- Machine learning infrastructure: Hosted machine learning models as model services leveraging internal MLops platform; designed and implemented feedback loops for and monitoring mechanism for online models.
- Data Engineering: Designed and implemented a scalable and robust ELT pipeline to process daily 40,000 production semiconductor wafer lots leveraging private cloud PaaS such as Apache Airflow, MariaDB, MinIO.
- Production System: Successfully refactored a legacy production system in collaboration with two Corp. IT divisions by integrating AI-enabled analytics into the document system, improving daily production operations for 110 production planners and manufacturing planners.
  - ✧ Business Impact: A more precise cycle-time-per-layer forecast supports production planners in envisioning potential output opportunity and designing optimal wafer start plan.

Cycle Time Forecast

- Data Analytics:
  - Applied machine learning methods such as SARIMA, gradient boosting regressors; deep learning methods such as encoder-decoder based recurrent neural networks, FEDformer in time series predictive modeling tasks.
  - Applied time-series based data augmentation techniques such as Fast Fourier Transform and pretraining, resulting in an improvement of 16.21% of MAPE in comparison to the legacy solution production control planners relied on.
- Data Governance:
  - Transformed large scale Fab and tool group data from various IT systems to 20 business domain data subjects that benefit to all business units of TSMC.
  - Designed data science development infrastructure with IT to benefit data scientists from accessing production zone data at development zone.
- ✧ Business Values:
  1. Shattered organization silos to leverage data from other divisions to build up data infrastructure and machine learning infrastructure for digital transformation.
  2. A more precise cycle time forecast improves delivery efficiency of TSMC products.

Cloud Architect **TSMC** **Nov, 2021 – Jan, 2023**

Enterprise-scaled Hybrid Cloud Foundation

- Designed IAM Roles and Responsibilities with AWS SSO Permission Sets.
- Applied Hub and Spoke topology to scale and manage hybrid cloud network infrastructure.
- Designed and implemented cross-account and cross-region security services logging with EventBridge, CloudWatch Log & Kinesis Data Firehose.
- Strong familiarity with Security standards on public cloud such as Control Tower Guardrails,

Organization Service Control Policies and Security Hub CIS & FSBP.

- ✧ Business Values:

1. Spearheaded the design and implementation of AWS hybrid cloud foundation at TSMC.
2. Upcoming projects benefit security control, scalable hybrid network infrastructure and predefined Permission Sets from this robust foundation.

Auto-remediation on Security Hub Findings

- Implemented auto-remediation and detective control on AWS Security Hub findings with CloudFormation StackSets, custom Config Rules and custom Systems Manager Document.
- Developed serverless Lambda function for custom Config Rule and custom Systems Manager Document.
- ✧ Business Value: Improved security posture across the AWS Organization, boosting the default Security Score in AWS Security Hub. New accounts automatically inherit auto-remediation mechanisms, reducing manual remediation efforts for non-compliant resources.

SAP HANA Migration

- Designed and implemented three-tier architecture for SAP HANA enterprise servers.
- Imported on-premise VMDK as AMIs and restore them as EC2 instances.
- ✧ Business Value: Migrated large-scale on-prem SAP servers to public cloud to save compute and storage costs.

Log Centralization Platform

- Designed and implemented an ELT pipeline to serve Terabytes of daily raw logs with Function App, Event Grid and Data Explorer.
- ✧ Business Value: Consolidated security logs from various platforms and saved 80% of storage costs.

Solutions Architect **Amazon Web Services** **June, 2020 – Feb, 2021**

- Delivered scalable cloud solutions to enterprise customers using Amazon Lookout for Vision, Amazon SageMaker, and Storage Gateway with S3.
- Led AWS Immersion Day & hands-on labs at customer sites to demonstrate AWS services and best practices.
- Developed an object detection solution using AWS DeepLens, Kinesis Video Streams, and AWS IoT:
  - Deployed MXNET SSD ResNet50 on AWS DeepLens with AWS Greengrass SDK.
  - Delivered inference results via AWS IoT with MQTT.
  - Streamed near real-time camera feed using Kinesis Video Stream Producer on AWS Console.
- Implemented facial recognition with Amazon Kinesis Data Streams and Rekognition Video:
  - Configured Rekognition Video as a Kinesis Data Stream consumer.
  - Built and indexed Rekognition collections with known face images.
- Applied Amazon Forecast DeepAR+ algorithm for demand forecasting:
  - Achieved improved results (lower MAD and MAPE) compared to LSTM.

Institute of Mathematics Academia

Research Assistant **Sinica** **Summer 2019**

- Courses on: Martingale pricing to derivatives contracts, stochastic process, decentralized Finance
- Apply Long short-term memory (LSTM), a Recurrent Neural Network, on time-serialized data

EDUCATION

Taipei, TW **National Taiwan University** **Fall 2019 – 2021**

- M.S in Information Management, GPA: 4.1/ 4.3
- Coursework: Machine Learning Foundation, Financial Time Series, Information Retrieval and Text Mining, Cognitive Computing, Statistical Learning, Software Development Methods, Information Management
- Built an Desktop App AlphaCrawler with Python and Kivy framework
    - Served to crawl raw text data from a crowd-sourced content service for financial markets and write to MySQL database server
    - Optimize poor system performance occurred from heavy concurrent transactions by introducing InnoDB engine Update Lock
- Ranked 1st among 300 teams in the NLP Money Laundering Tournament 2020 hosted by E.Sun Bank:
    - Led a team of 8 members, coordinating roles in data crawling, augmentation, labeling, model training, and inference over two months.
    - Completed two ML tasks: Binary classification of news as money laundering-related or not, and Named Entity Recognition (NER) to identify money laundering suspects.
    - Achieved 0.993 accuracy in binary classification and 0.915 F1-score in NER using Chinese XLNet-600.
- Ranked 1st place among 20 teams at Statistical Learning (Summer, 2020) Final Project
    - Investigated machine learning fairness in financial credit risk modeling by performing a binary classification to predict loan repayment capability.
    - Leveraged IBM's AIF260 to assess and ensure fairness in the model's predictions.
    - Binary Classification task: https://andwct.github.io/machine_learning_project/credit-risk-modeling-classification/
    - AI Fairness exploration: https://andwct.github.io/machine_learning_project/ai_fairness/
- Master thesis: Image Reconstruction from Hierarchical Representations of Human Visual Cortex Activities

CERTIFICATES

- AWS Machine Learning Specialist
- AWS Solutions Architect Associate
- IELTS 7.5

Techstacks

- **Programming Languages:** Python, Golang, Perl, R, Java, C++
- **ML/DL Frameworks:** MXNet, Pytorch, Caffe, Tensorflow, Keras, NumPy, Pandas, Scikit-Learn, XGBoost, LightGBM, HuggingFace
- **Cloud Platforms:** AWS, Azure, GCP
- **Infrastructure as code:** Hashicorp Terraform, Azure DevOps, AWS CloudFormation Stacks & Stacksets, AWS CDK