

# Nachai Limsettho

Senior Data Scientist (NLP specialist)

Nationality: Thai  
Birthday: 12 August 1988  
Mobile: +65 93234775  
Email: [nachailim@gmail.com](mailto:nachailim@gmail.com)  
Address: 140 Hillview Ave,  
Singapore 669600

## PROFESSIONAL SUMMARY

With over a decade of experience as a seasoned Data Scientist, I offer extensive expertise, particularly in Natural Language Processing (NLP), complemented by a strong technical background and hands-on involvement in data handling, manipulation, and analysis across dynamic sectors such as media, insurance, telecom (Telco), and e-payments.

Throughout my career, I've dedicated myself to research and development in data mining and machine learning, resulting in a specialization in cutting-edge NLP techniques and machine learning algorithms. My practical experience includes proficiency in advanced deep-learning techniques like LLM, DNN, CNN, and Word2Vec. Additionally, I am adept at navigating cloud environments, particularly AWS, as well as working on private servers, making me a versatile and knowledgeable asset in the field of data science and machine learning.

**Project highlight:** Sentiment Analysis model, Recommendation System, Credit Scoring, Moderation Projects through AWS, Customer Scores from Transaction Behaviours

## SKILLS

**Programming:** Python, PySpark, SQL, C#  
**Cloud:** Amazon Web Services (AWS), S3, Redshift, SageMaker, Rekognition, Textract  
**Deep Learning:** LLM, CNN, LSTM, DNN, RNN, Transformer, Auto-Encoder, etc.  
**Data Mining:** Classification, Clustering, Ensemble, Imbalanced dataset, etc.  
**NLP:** LLM, Sentiment Analysis, Word2vec, Tokenization, etc.  
**Collaboration Tools:** GitLab, GitHub, Notion, and JIRA  
**Data Visualization:** Tableau, Pyplot, and Seaborn  
**Languages:** English, Thai, Japanese (Basic)  
**Generals:** Project Management, Interdivisional Collaboration. Presentation Skills, Well organized, Good at teamwork

## AWARD AND ACHIEVEMENTS

**2017** Invited speaker at Monetizing Big Data in Telecoms World Summit  
Title: Mining for High-End Real Estate Customers - A Case Study of Utilizing Telecom Database for Potential Customers Discovery  
Venue: Singapore, Singapore  
**2013-2016** Japanese Government Scholarship (Mongbukagakusho)  
**2010-2012** Scholarship from the Department of Computer Engineering, KU

## WORK HISTORY

### Senior Data Scientist · TipTip Network PTE. LTD. (Singapore)

2022 - Present

- ❖ Led, Designed, and implemented a comprehensive **Digital Content and Creator recommendation system on AWS**, incorporating both behavioral and encoded textual features. This system efficiently ranks items and creators for customers based on their preferences.
- ❖ Engineered and **implemented an automated eKYC system** utilizing AWS Rekognition & Textract technologies. This system adeptly recognizes identity information from ID cards and extracts textual data, facilitating seamless integration with bank account details for user identity verification processes
- ❖ Designed and **deployed a video moderation system** that automates the identification of offensive imagery within uploaded videos. Through a customized approach, this solution significantly reduced Rekognition API costs by 80% while maintaining high detection performance standards.
- ❖ Developed and **implemented a text moderation system** employing a lexicon-based methodology. This system automates the detection of offensive language within uploaded PDF files and accompanying descriptions, while also scanning cover and inserted images for inappropriate content.
- ❖ **Implemented a system for intellectual property protection**, utilizing fuzzy logic and file information to detect similar PDFs. This contributes to safeguarding intellectual property through effective and efficient detection methods.

### Data Scientist · OVO (PT Visionet Internasional) (Singapore)

2019 - 2022

- ❖ Led the development and **implementation of multiple customer score models**, leveraging extensive transactional data from a user base exceeding fifteen million. These models translated complex customer behaviors into actionable numerical insights, with many seamlessly integrated into production systems and enthusiastically adopted by business stakeholders.
- ❖ **Created and executed a sentiment analysis model** intricately fused with NLP and a topic model, enabling comprehensive analysis of customer feedback sourced from diverse external channels. The resulting insights empowered local teams to swiftly and accurately pinpoint customer pain points, facilitating proactive resolution strategies.
- ❖ **Developed and engineered a feature store** containing over three hundred advanced features for use across diverse projects and models.
- ❖ Collaborated with Jakarta-based teams to address various challenges and devise business solutions tailored to specific requirements

### Senior Executive, Data Analytics · Allianz SE (Singapore)

2018-2019

- ❖ Collaborated with the risk team to uncover crucial insights for informed decision-making. Additionally, developed a corresponding Tableau dashboard for business users to visually assess SEA OEs performance.
- ❖ Led the management and **implementation of a Call Volume Prediction model** utilizing Recurrent Neural Networks (RNN) to optimize staff allocation procedures for call centers in Thailand.

- ❖ Led multiple data science projects for Thailand OE, exploring issues related to claims, policy data, and various other aspects to drive effective solutions.
- ❖ Analyzed and visualized diverse projects in response to requests from international OEs across Southeast Asia.
- ❖ Collaborated with various local functional teams to implement models and track their outcomes.

#### **Data Scientist** · *True Corporation Public Company Limited. (Thailand)* 2016-2018

- ❖ Evaluated and reviewed the technological aspects of foreign and local companies
- ❖ Provided technical consultations to top management to facilitate informed decision-making.
- ❖ Created a prototype classification model utilizing a telecom database and neural network to recommend potential customers for a non-telecom partner.
- ❖ **Developed models to calculate affinity scores**, distinguishing customer lifestyles based on their browsing history and application usage.
- ❖ Analyzed and visualized characteristics of privileged customers for customer retention, leveraging correlations for insightful findings.
- ❖ Prepared and processed telecom data to answer various business needs.

## EDUCATION

2013-2016

Nara Institute of Science and Technology, Nara, Japan

**Doctor of Philosophy in Computer Engineering**

Laboratory of Software Engineering

2010-2012

Kasetsart University, Bangkok, Thailand

**Master in Computer Engineering (GPA: 3.90)**

Data Analysis and Knowledge Discovery Laboratory (DAKDL)

2006-2010

Kasetsart University, Bangkok, Thailand

**Bachelor in Computer Engineering (GPA: 3.10)**

## SELECTED PUBLICATION

### **Data Mining & Machine Learning:**

- 2014**      **“Integrating Weight with Ensemble to Handle Changes in Class Distribution”** In 10th International Conference on Machine Learning and Data Mining in Pattern Recognition, 2014
- More Extension of the previous work, this time each model in the selected set is weighted based on how much they are difference from the estimated class distribution*
- 2011**      **“Handling Concept Drift via Ensemble and Class Distribution Estimation Technique”** In 7th International Conference on Advanced Data Mining (ADMA 2011), 2011
- Extension of the previous work, more sophisticated ensemble frameworks are proposed to select only a set of models that are suitable for the current dataset distribution.*

- 2011**      **“Quantification with Ensemble Method for Handling Change in Class Distribution”** In the 2011 International Computer Science and Engineering Conference (ICSEC 2011), 2011  
*Develop an ensemble of quantification classifiers to handle Concept Drift, a problem that the class distribution of the target dataset is change over time.*

### Software Engineering and NLP:

- 2018**      **“Cross-project defect prediction using class distribution estimation and oversampling”** Information & Software Technology, Volume 100, 2018  
*More Extension of the previous work, improving the imbalanced dataset issue in defect prediction with optimized oversampling approach*
- 2016**      **“Unsupervised Bug Report Categorization using Clustering and Labelling Algorithm”** In the International Journal of Software Engineering and Knowledge Engineering (IJSEKE), 2016  
*Extension of the previous work, including an automatic method to label the categorized bug reports*
- 2016**      **“A Strategy to Determine When to Stop Using Automatic Bug Localization”** In 40th IEEE Annual Computer Software and Applications Conference (COMPSAC), 2016  
*Study of how can we minimize the efforts spent in bug localization*
- 2014**      **“Automatic Unsupervised Bug Report Categorization”** In 2014 6th International Workshop on Empirical Software Engineering in Practice (IWESEP), 2014  
*Proposed a framework to automatically categorize bug reports without any labeled dataset using clustering and topic modeling.*
- 2014**      **“Comparing hierarchical Dirichlet process with latent Dirichlet allocation in bug report multiclass classification”** In 15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2014  
*A study to compare the performance of two well-known topic models in bug report categorization, these models are utilized to extract topic features from bug reports in their natural language form.*

### SELECTED CERTIFICATE

- ❖ (IBM) Generative AI: Foundation Models and Platforms  
<https://coursera.org/share/a27d4dec4efb8b085a67b287fe70f1fc>
- ❖ (IBM) Generative AI for Data Scientists Specialization  
<https://coursera.org/share/2e99698ce68f8c9e5304bed3aa58dc7f>
- ❖ (IBM) Generative AI Fundamentals Specialization  
<https://coursera.org/share/6cd52dc95bedfa8937efd5f7b3c0ad3d>
- ❖ (Duke University) Python Essentials for MLOps  
<https://coursera.org/share/d776f629ae29569f7ee2913a2f4d7c3d>
- ❖ (Duke University) DevOps, DataOps, MLOps  
<https://coursera.org/share/bdee76a4d33c3a523f4354895bd716af>
- ❖ (University of Glasgow) Generative Pre-trained Transformers (GPT)  
<https://coursera.org/share/ef69baf22f2ad1b87385298710221f1e>