

CHEA ROTH

DATA SCIENCE AND MACHINE LEARNING

October 22, 2025

+885-97 4865 856

<https://github.com/Rotha-101>

chea-rotha.itc.edu@gmail.com

Toul Kork, Phnom Penh, Cambodia

<https://www.linkedin.com/in/chea-rotha-44268b2a5/>

SUMMARY

I am CHEA ROTH, a Data Scientist with an Engineering Degree in Data Science from the Institute of Technology of Cambodia (ITC). I am proficient in leveraging Machine Learning (ML), time series analysis, and data visualization to drive data-driven solutions. Eager to apply strong academic foundations and hands-on project experience in a professional environment to deliver innovative insights.

EDUCATION BACKGROUND

2020 – 2025

INSTITUTE OF TECHNOLOGY OF CAMBODIA (ITC)

ENGINEERING DEGREE IN DATA SCIENCE

- Major in Apply Mathematics and Statistics
- Thesis on "*Analysis and Forecasting on Inflation in Cambodia*"

2019 - 2022

AII LANGUAGE CENTER

- English
- French (Basic)

PROFESSIONAL EXPERIENCE

SCHOOL PROJECTS

LAPTOP PRICE PREDICTION

Role: Data Analyst & Model Developer

Tools: Python, Pandas, Scikit-learn, BeautifulSoup, Matplotlib, Seaborn

- Performed data cleaning, exploratory data analysis (EDA), and feature engineering.
- Applied multiple regression models including Linear Regression, Random Forest, and Support Vector Regressor (SVR).
- Achieved high prediction accuracy with the Support Vector Regressor model.

FORECASTING TOURISM IN CAMBODIA

Role: Data Analyst / Time Series Forecaster

Tools & Skills: Python (Pandas, NumPy, Matplotlib), ARIMA, SARIMA, Prophet, LSTM, Time Series Modeling

- Collected and preprocessed tourism data for time series analysis.
- Evaluated model performance using Mean Squared Error (MSE) and residual diagnostics.
- Identified seasonal patterns and recovery trends post-COVID-19.
- Recommended data-driven insights for tourism recovery planning and policy-making.

PROFESSIONAL SKILLS

- Python Programming (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn)
- Data Analysis & Visualization (Power BI, Excel)
- Machine Learning (Regression, Classification, Clustering)
- Deep learning
- SQL & Database Management
- Statistics & Probability
- Microsoft Office Suite (Word, PowerPoint, Excel)
- Research & Report Writing
- Teamwork & Communication

INSACART MARKET BASKET ANALYSIS

Role: Data Analyst

Tools & Skills: Python (Pandas, NumPy, Matplotlib, Seaborn), Time Series Analysis

- Performed time series and customer behavior analysis on an e-commerce dataset inspired by Instacart.
- Conducted data cleaning, preprocessing, and visualization to identify shopping patterns and product demand trends.
- Applied RFM analysis to segment customers by recency, frequency, and spending behavior.
- Delivered actionable insights to improve marketing, inventory planning, and customer retention strategies.

HOTEL RESERVATION SYSTEM DATABASE

Role: Database Designer & Developer

Tools & Skills: MySQL, SQL, ERD Design, Relational Database Modeling

- Designed and implemented a database system to manage hotel operations such as room booking, client management, and staff scheduling.
- Built ER diagrams and relational schemas for key entities Client, Staff, Room, Service, Floor, Booking.
- Applied data normalization, primary and foreign key constraints to ensure data integrity.
- Supported receptionist workflows including booking verification and payment confirmation.

CROP YIELD PREDICTION & RECOMMENDATION

Role: Data Analyst

Tools & Skills: Python (Pandas, NumPy, Matplotlib, Seaborn), Time Series Analysis, RFM Segmentation

- Laptop Price Prediction (Machine Learning): Developed models to predict laptop prices using regression algorithms.
- Crop Yield Prediction & Recommendation: Analyzed agricultural data to provide actionable insights for farmers.
- Movie Recommendation System (Machine Learning): Built and evaluated a linear regression-based movie recommender.

DEVELOPMENT OF SPECIALIZED TOKENIZER FOR KHMER LARGE LANGUAGE MODEL (LLM)

Role: Data Analyst

- Engineered a custom SentencePiece tokenizer (7,000 tokens) to minimize linguistic bias for Khmer.
- Reduced token sequence count by over 90% (from 42,914 to 3,617 on sample data) compared to LLaMA2, significantly boosting processing efficiency and model accuracy.
- Managed Khmer Corpus Creation and utilized LoRA for efficient model fine-tuning.

SUNRISE INSTITUTE - JAN 2024 - FEB 2025

Instructor – Data Science

- Taught Fundamentals of Data Science, covering topics such as data collection, cleaning, analysis, and model interpretation.
- Provided hands-on training in Data Visualization using Power BI and Python (Matplotlib, Seaborn, Pandas).
- Guided students in building interactive dashboards, analyzing real-world datasets, and communicating insights effectively.
- Designed and delivered practical exercises to help learners apply statistical and analytical concepts in business contexts.
- Mentored students on best practices in data storytelling, visualization design, and data-driven decision-making.

MINISTRY OF PLANNING - FEB 2025 - PRESENT

Role: Data Analyst & Data Scientist

Tools: Python, Pandas, Scikit-learn, BeautifulSoup, Matplotlib, Seaborn

- Developed and deployed a hybrid inflation forecasting framework using Python, combining Machine Learning (XGBoost, LSTM) and SARIMAX to produce national economic projections.
- Led the end-to-end data pipeline for critical economic indicators, encompassing data cleaning, feature engineering, and stationarity testing.
- Provided evidence-based policy analysis to senior officials on the influence of sectoral and global factors (e.g., oil prices, gold prices) on domestic inflation trends.
- Communicated complex analytical findings through tailored reports and dashboards to support proactive policy formulation.

- Python Programming (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn)
- Data Analysis & Visualization (Power BI, Excel)
- Machine Learning (Regression, Classification, Clustering)
- SQL & Database Management
- Statistics & Probability
- Microsoft Office Suite (Word, PowerPoint, Excel)
- Research & Report Writing
- Teamwork & Communication

EDUCATION BACKGROUND

PHAUK SOKKHEY, Ph.D

Position: Deputy head of department of applied mathematics and statistics (AMS) (អនុប្រធានដេប៉ាតឺម៉ង់គណិតវិទ្យាអនុវត្ត និងស្ថិតិ) and Head of Master Program of Data Science

Phone: 012 939 310
Email : phauk.sokkhey@itc.edu.kh

Mr. Pakrina Long

Position: IT Lectureer in DATA Science and Institute of Technology Cambodia

Phone: 010 64 44 96
Email : long.pakrigna@itc.edu.kh