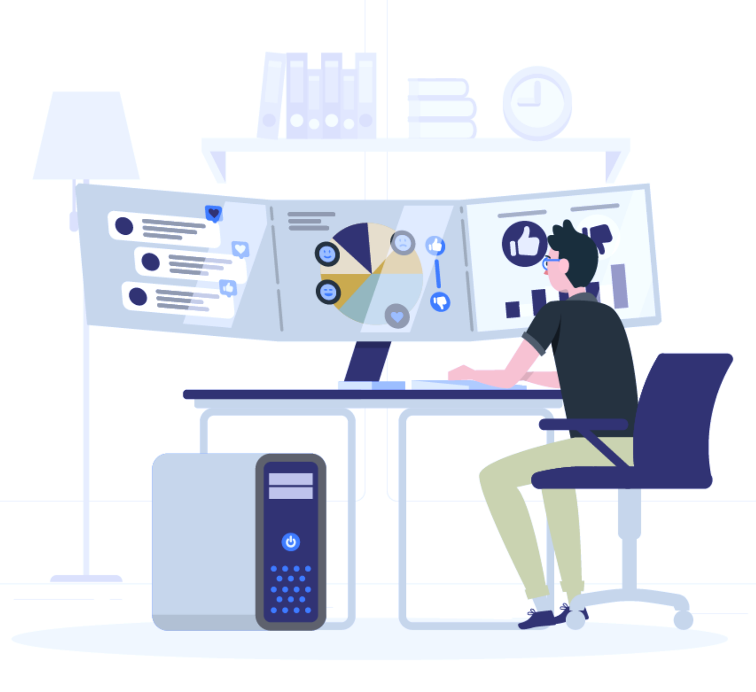
Leveraging Data Science Techniques for Predictive Modeling of Global Skill Penetration Rates

Student: Roatny NUON (M080105)

M-AMS

Aug.2023

Table of contents



01

Introduction

M-AMS

Background

In today's world, where decisions are increasingly guided by data-driven insights, this research search through into the practical applications of Data Science. Through techniques like data preprocessing, predictive modeling, and interpretation, the study demonstrates how Data Science turns intricate data into actionable wisdom. In parallel, the research aligns seamlessly with Cambodia's developmental vision and mirrors its Sustainable Development Goals (SDGs). Specifically, it contributes to the realization of SDG 9 – Industry, Innovation, and Infrastructure – by predicting skill usage trends across industries, propelling the adoption of innovative technologies. Additionally, it holds the potential to fortify SDG 4 – Quality Education – by spotlighting how Data Science methods amplify workforce training strategies.

Problem Statement

Shifting Skill Demands

- Rapid technological changes

Precision Gap in Workforce Planning

- Lack accuracy and agility to predict skill penetration rates effectively

The Need for Data-Driven Precision

- Data-driven approaches to forecast global skill penetration trends

Impower & Contributed

- Decision Maker
- Workforce Planner

Research Objectives

1. Apply Data Science techniques to preprocess, transform, and analyze text-based World Bank data for predictive modeling of global skill penetration rates.
2. Develop and optimize predictive models using regression algorithms, showcasing the practical integration of machine learning in addressing real-world challenges.
3. Explore correlations between skill penetration trends, economic indicators, and SDG priorities, specifically relevant to Cambodia's development context.
4. Investigate how predictive modeling insights can contribute to the Cambodia Digital Economy and Society Policy Framework's objectives for workforce development and digital transformation.

Scope and Limitation

SCOPE

- The research will utilize the "public_use-skill-penetration" dataset from Word Bank, spanning the years 2015-2019, to predict skill penetration trends using Data Science Techniques.
- Global Perspective: The research will target industries on a global scale, encompassing diverse sectors and regions to capture broad trends in skill demand.

Scope and Limitation

LIMITATION

1.Generalization: The global scope might limit the model's ability to capture nuances specific to individual countries or regions, potentially impacting prediction accuracy at a local level.

2.Assumption of Stationarity: The research assumes that skill penetration trends remain relatively stable over the study period, which might not account for abrupt shifts caused by unforeseen events.

3.External Factors: Economic, political, or social factors that aren't explicitly captured in the dataset might influence skill penetration rates, affecting the model's predictive capabilities.

02

Literature Review

M-AMS

Literature Reviews-4 papers

Title/Date	Objective	Methodology	Result/ Conclusion
"Developing a Machine Learning-based Job Forecasting and Trend Analysis System for Predicting Future Job Markets Using Historical Data." SENTHURVELAUTHAM, S. and SENANAYAKE, N. 2023	Address the gap in understanding of job market trends in the IT field, specifically focusing on roles and skills factors Building Auto-aggressive prediction that predict for various roles and technologies for at least the next 12 months	Data collection: Web scraping, Manually collecting data & Govern. Dataset > 522,180 job listings from the past 24 months in the software industry IF-IDF, ARIMA, LSTM, GRU, SARIMA, Prophet, DeepAR bi-directional LSTM model	Accuracy of 95.71%.
Forecasting unemployment with Google Trends: age, gender and digital divide Rodrigo Mulero and Alfredo Garcia-Hiernaux, 2022	This paper uses time series of job search queries from Google Trends to predict the unemployment in Spain Paper exploits GT time-series data from 2004 to 2018, on a collection of more than 170 search-related items, to predict unemployment figures	Jan. 2004 to Sep. 2018, for a total of 177 monthly observations The data are disaggregated by age and gender groups ARIMA & SARIMA gave result similar. Evaluate model: AIC,RMSE,	RMSE is 14.5% for the youngest unemployed

Literature Reviews-4 papers

Title/Date	Objective	Methodology	Result/Conclusion
Dr. Padmaja Pulicherla et al 2019 J. Phys.: Conf. Ser. 1228 012056 “Job Shifting Prediction and Analysis Using Machine Learning”	Designed a predictive model to anticipate the chances of an employee leaving the job	18K data from Analytics Vidhya site Random Forest , XGBoost , CatBoost , LightGBM	CatBoost is the best model with 0.7%
T. V. Huynh, K. V. Nguyen, N. L. T. Nguyen, and A. G. T. Nguyen, "Job Prediction: From Deep Neural Network Models to Applications," 2020	Studying the job prediction using different deep neural network models including Text CNN,Bi-GRU-LSTM-CNN, and Bi-GRU-CNN with various pre-trained word embeddings on the IT job dataset. **Input:** Given an IT job description collected from the online finding-job sites **Output:** A predicted job title for this description.	Dataset 10,000 distinct job descriptions, from the online finding-job sites w/ 25 different types of IT-related job. 4 deep neural network models used: TextCNN, two combination models (Bi-GRU-CNN and Bi-GRU-LSTM-CNN) and ensemble mode	In the Bi-GRU-CNN + Glove model, the class "IT Consultant" fot largest accuracy of 91.43% and the class "Data Architect” got lowest accuracy of 44.19%

03

Data Collection

M-AMS

Data Background

Type: Secondary Data

LinkedIn Data World Bank.ORG

SIZE

20781 x 7

Includes information about skill group penetration rates across various industries and years.

Features

year, skill group category, skill group name, ISIC section index, ISIC section name, industry name, and skill group penetration rate.

WORLD BANK GROUP
THE WORLD BANK IBRD • IDA | IFC International Finance Corporation | MIGA Multilateral Investment Guarantee Agency

LinkedIn

Digital Data for Development

Skill Penetration Dataset

Context

The LinkedIn and World Bank Group collaboration is a prime example of how technology companies can work with development institutions to bring new data and insights to developing countries to address pressing development challenges. The opportunities and challenges presented by the global economy requires the public and private sectors to join forces, share information, share resources, and work towards a common vision to make meaningful, positive and scalable impact.

The dataset underlies the metrics presented on the interactive dashboard of the World Bank Group – LinkedIn partnership. The metrics and data are a product of the partnership and cover industry employment, skill, and migration metrics for over 100 countries. Specifically, the data covers 4 metrics: 1) Industry Employment Shifts, 2) Talent Migration, 3) Industry Skills Needs, and 4) Skill Penetration. LinkedIn and the World Bank Group plan to refresh the data annually at a minimum.

This file contains Skill Penetration data

Caveats

The Skill Penetration metric looks at how many skills from each of LinkedIn's skill groups (see "Notes" tab) appear among the top 30 skills for each occupation in an industry. For example, if 3 of 30 skills for Data Scientists in the Information Services industry fall into the Artificial Intelligence skill group, Artificial Intelligence has a 10% penetration for Data Scientists in Information Services. These penetration rates are averaged across occupations to derive the industry averages reported. It is likely this metric is best at capturing skill penetration across tradable and knowledge-intensive sectors. For example, it may under-estimate the adoption of AI in Manufacturing, since LinkedIn members are less likely to be in this sector compared to others.

For more information on the World-Bank Group - LinkedIn Partnership, please visit linkedindata.worldbank.org

https://linkedindata.worldbank.org/data

Data Table

	A	B	C	D	E	F	G	H
1	year	skill_group_c	skill_group_r	isic_section_	isic_section_	industry_nan	skill_group_penetration_rate	
2	2015	Business Skil	Accounts Pay	M	Professional	Accounting	0,00719	
3	2015	Business Skil	Accounts Pay	M	Professional	Law Practice	0,00244	
4	2015	Business Skil	Accounts Pay	M	Professional	Executive Of	0,00222	
5	2015	Business Skil	Accounts Pay	C	Manufacturi	Packaging &	0,00132	
6	2015	Business Skil	Accounts Pay	B	Mining and c	Oil & Energy	0,00132	
7	2015	Business Skil	Accounts Pay	C	Manufacturi	Printing	0,00128	
8	2015	Business Skil	Accounts Pay	C	Manufacturi	Machinery	0,00112	
9	2015	Business Skil	Accounts Pay	K	Financial and	Venture Capi	0,00102	
10	2015	Business Skil	Accounts Pay	C	Manufacturi	Plastics	9,40E-04	
11	2015	Business Skil	Accounts Pay	M	Professional	Legal Service	9,20E-04	
12	2015	Business Skil	Accounts Pay	C	Manufacturi	Textiles	8,30E-04	
13	2015	Business Skil	Accounts Pay	B	Mining and c	Mining & Me	8,10E-04	
14	2015	Business Skil	Accounts Pay	M	Professional	Mechanical (	7,70E-04	
15	2015	Business Skil	Accounts Pay	C	Manufacturi	Electrical & E	7,50E-04	
16	2015	Business Skil	Accounts Pay	C	Manufacturi	Food Product	7,50E-04	
17	2015	Business Skil	Accounts Pay	M	Professional	Environment	7,30E-04	
18	2015	Business Skil	Accounts Pay	C	Manufacturi	Automotive	7,10E-04	
19	2015	Business Skil	Accounts Pay	J	Information	Information	6,60E-04	
20	2015	Business Skil	Accounts Pay	M	Professional	Architecture	6,10E-04	
21	2015	Business Skil	Accounts Pay	C	Manufacturi	Industrial Au	6,00E-04	
22	2015	Business Skil	Accounts Pay	C	Manufacturi	Chemicals	5,70E-04	
23	2015	Business Skil	Accounts Pay	J	Information	Telecommur	5,50E-04	
24	2015	Business Skil	Accounts Pay	J	Information	Computer So	5,30E-04	
25	2015	Business Skil	Accounts Pay	M	Professional	Outsourcing/	5,20E-04	
26	2015	Business Skil	Accounts Pay	M	Professional	Marketing &	5,10E-04	
27	2015	Business Skil	Accounts Pay	M	Professional	Biotechnolog	5,00E-04	
28	2015	Business Skil	Accounts Pay	R	Arts, enterta	Health, Well	4,90E-04	
29	2015	Business Skil	Accounts Pay	K	Financial and	Financial Ser	4,80E-04	

20750	2019	Tech Skills	Web Develop	J	Information	Broadcast M	0,00453
20751	2019	Tech Skills	Web Develop	C	Manufacturi	Industrial Au	0,00435
20752	2019	Tech Skills	Web Develop	K	Financial and	Investment M	0,00409
20753	2019	Tech Skills	Web Develop	M	Professional	Photography	0,00406
20754	2019	Tech Skills	Web Develop	R	Arts, enterta	Entertainme	0,00405
20755	2019	Tech Skills	Web Develop	C	Manufacturi	Electrical & E	0,00401
20756	2019	Tech Skills	Web Develop	R	Arts, enterta	Sports	0,00398
20757	2019	Tech Skills	Web Develop	C	Manufacturi	Renewables	0,00356
20758	2019	Tech Skills	Web Develop	M	Professional	Accounting	0,00354
20759	2019	Tech Skills	Web Develop	M	Professional	Legal Service	0,00353
20760	2019	Tech Skills	Web Develop	M	Professional	Environment	0,00311
20761	2019	Tech Skills	Web Develop	M	Professional	Biotechnolog	0,00309
20762	2019	Tech Skills	Web Develop	M	Professional	Architecture	0,00261
20763	2019	Tech Skills	Web Develop	C	Manufacturi	Aviation & A	0,00261
20764	2019	Tech Skills	Web Develop	C	Manufacturi	Automotive	0,00253
20765	2019	Tech Skills	Web Develop	M	Professional	Law Practice	0,00247
20766	2019	Tech Skills	Web Develop	J	Information	Semiconduct	0,0024
20767	2019	Tech Skills	Web Develop	M	Professional	Mechanical (	0,00226
20768	2019	Tech Skills	Web Develop	C	Manufacturi	Pharmaceuti	0,00198
20769	2019	Tech Skills	Web Develop	J	Information	Motion Pictu	0,00194
20770	2019	Tech Skills	Web Develop	C	Manufacturi	Textiles	0,00192
20771	2019	Tech Skills	Web Develop	M	Professional	Executive Of	0,00184
20772	2019	Tech Skills	Web Develop	R	Arts, enterta	Museums &	0,00172
20773	2019	Tech Skills	Web Develop	C	Manufacturi	Machinery	0,0017
20774	2019	Tech Skills	Web Develop	B	Mining and c	Oil & Energy	0,00168
20775	2019	Tech Skills	Web Develop	C	Manufacturi	Packaging &	0,00159
20776	2019	Tech Skills	Web Develop	B	Mining and c	Mining & Me	0,00147
20777	2019	Tech Skills	Web Develop	C	Manufacturi	Food Product	0,00144
20778	2019	Tech Skills	Web Develop	C	Manufacturi	Chemicals	0,00139

04

Methodology

M-AMS

How to achieve Predictive Task

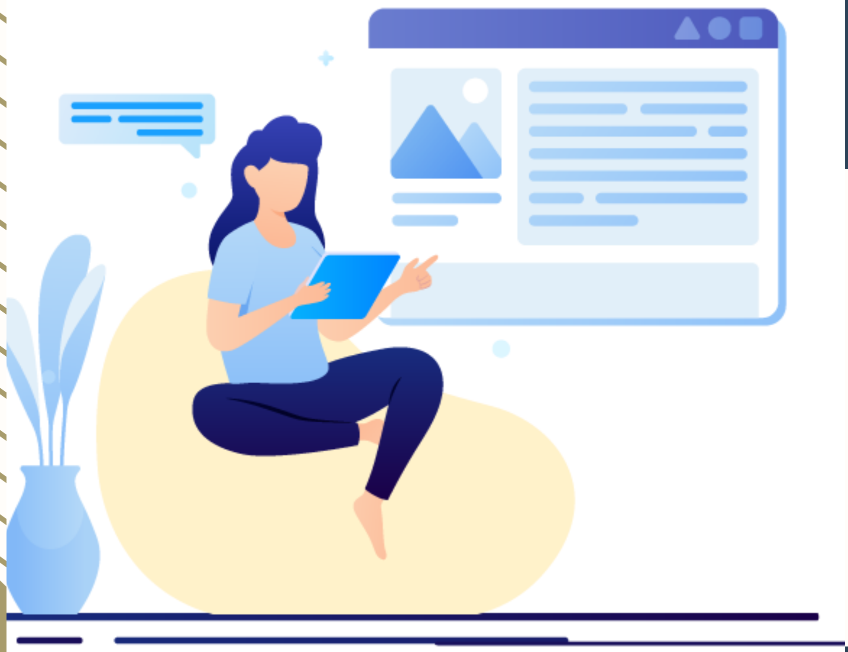
Preprocessing
- Convert data using hybrid techniques one-hot encoding and NLP
- Apply embedding(word2vec, Glove)

Feature Engineering
- Use the word embeddings as additional input features alongside the time series data

Data Preparing_TS
- Capture trend in data

Train Model LSTM vs XGBoost
- Prepare input data for each Model
- Train
- Will Test on more Algorithms

Evaluate & Interpretation
- Evaluate
- Compare
- Interpret result to extract insight into future skill trend



05

Expected Result

M-AMS

Expected Result

Based on Final Model after Model Evaluation

Numerical Prediction

Obtain numerical predictions of future skill penetration rates over time intervals.

Insights into Skill Trends

The model will provide insights into how skill penetration rates are influenced by both historical trends

Interpretable Results

Interpretable than traditional methods

Actionable insights for workforce planning & decision-making related to skill development and industry trends.

References

SENTHURVELAUTHAM, S. and SENANAYAKE, N. 2023 "Developing a Machine Learning-based Job Forecasting and Trend Analysis System for Predicting Future Job Markets Using Historical Data”

Mulero and Alfredo Garcia-Hiernaux, 2022, Forecasting unemployment with Google Trends: age, gender and digital divide

Dr. Padmaja Pulicherla et al 2019 J. Phys.: Conf. Ser. 1228 012056“Job Shifting Prediction and Analysis Using Machine Learning”

T. V. Huynh, K. V. Nguyen, N. L. T. Nguyen, and A. G. T. Nguyen, "Job Prediction: From Deep Neural Network Models to Applications,” 2020

Data source: https://linkedindata.worldbank.org/data

Thank You!

M-AMS