Jobs Prediction Model based on Myanmar Labor Force Survey by using Artificial Neural Network

Su Nandar
University of Information Technology
Yangon, Myanmar
sunandar@uit.edu.mm

Myint Myint Lwin
University of Information Technology
Yangon, Myanmar
myintmyintlwin@uit.edu.mm

Abstract*—Nowadays, job prediction is an essential task for every organization because of the influence of job fitting or job suitability. Job fit is a concept that refers to how well an employee is suited for his or her position and which can have a positive impact on the organization's morale. Therefore, the job prediction model is needed for both employees and employers. In this paper, the job prediction model for Myanmar employment is proposed and implemented by using the Artificial Neural Network (ANN) and back-propagation algorithm. The proposed model learns the real dataset of the Myanmar Labor which is collected from the Department of Labor, Ministry of Labor, and Immigration. The International Standard Classification of Occupations (ISCO) defines ten main job classes. Therefore, the proposed model can predict these ten job classes. According to the experimental result, the proposed model achieved the best result with an accuracy of 97%. The prediction result of the proposed system can also help not only decision-makers who worked in the management level of private or government organizations but also job seekers.

Keywords—Jobs Prediction, Predictive Analysis, Artificial Neural Network, Back propagation

I. INTRODUCTION

Nowadays, job prediction is very popular and useful for people who are seeking the desired job. In Myanmar, finding a suitable job or carrier path is a challenge for employees and applicants. Job fitting makes the productivity of people much better and creates a win-win relationship. While graduates also want to work with their professional qualifications, employers also want to get the employees who fulfill with required qualifications. Graduates' employability continues as national issues due to the increasing number of graduates produced by education institutions yearly. Unemployment is a big challenge for our country, Myanmar, and recruitment programs are the key indicators of economic development level for national economies. The solutions are required for unemployment problems and accommodation of employee's qualifications for their jobs.

For developing countries, greater competitiveness and productive economy need to consider the job fitting as one condition for national economic development. Human resources must be improved by countries which are suffering from unemployment. Jobs prediction is an important utility for the countries and it is necessary to recruitment for people's educational background.

Many researchers try to improve society by applying information technology especially Artificial Intelligence (AI). Machine learning is an application of AI that affords systems for automatically enhancing the experience without being predetermined. Learning methods may be supervised, semi-supervised, or unsupervised. The artificial neural network (ANN) is the supervised machine learning model that mimics a biological system. The ANN can recognize patterns and trends that are too complex for humans.

Recently, ANN has been successfully applied to solve a variety of problems that involves transactions in decision making, classifying, predicting, and optimization. Artificial Neural Networks modeling is considered an effective analysis method for prediction and can well support the correlation of nonlinear time series in delay state space.

The information of all working people especially their respective education can be analyzed to predict future jobs. Therefore, ANN is applied in the proposed job prediction model based on Myanmar Labor Information. This study investigates the job class to recognize the future jobs available and users could know about what qualifications are needed for their desired jobs.

To achieve the desired result, this paper proposes the job prediction model for Myanmar by using an Artificial Neural Network (ANN) and it is built on a real dataset of Myanmar Labor for the year 2018 and 2019 from the Department of Labor, Ministry of Labor, Immigration, and Population. The proposed model can be applied in private or public companies to hire people who have specific jobs qualification. And then, the system's result supports the decision-maker in the academic sector of Myanmar because it predicts job categories based on education, experiences, monthly wages, and other factors.

II. RELATED WORK

Many researchers tried to build a job prediction model to help the job seekers and employers of the organization. There are many prediction models are proposed by applying various analysis methods and learning models.

Akku Geroge Saju [1] investigated the comparative analysis by using a Decision tree, Linear Regression, Decision Forest, and Artificial Neural Network (ANN) to predict future job opportunities in Trivandrum city, India using 600 rows of data from the Institute and 809 rows of Data from Agency. The experimental result described the accuracy of using ANN is better than the other three machine learning methods used.

Anifat Olawoyin, Yangiuin Chen [2] compared a 4-layer multi-layer perceptron model with Auto-Regressive Integrated Moving Average (ARIMA) model for predicting the future parking contravention in Winnipeg, Canada. Time series parking tickets 1.09 million transactions with 5 Attributes from 2010-2016 are used to achieve the result. Multi-layer perceptron prediction produced the best goodness of fit (R2 = 0.77) and lowest prediction error (RMSE = 0.099). The author also investigated the effect of adding more layers on the performance of a multilayer perceptron neural.

Kuo-Yan Wang, Hsien-Yu Shun [3] focused to propose applying Back Propagation Neural Networks (BPNN) in the Prediction of Management Associate Work Retention for Small and Medium Enterprises. The authors utilized the data of 100 management trainees using BPNN to determine the probability of retaining management associates. When 20 Management Associate program participants are used as a test sample, high-level accuracy was gained from the proposed Back-propagation model.

Nedhal A. Al-Saiyd, Amjad S. Al-Takrouri [4] has applied BPNN (Back Propagation Neural Network) to predict IT jobs. This study used data set gained by interviewing as questionnaires included 35 input factors for 50 graduate students that cover IT job skills. The results from the Artificial neural network model in the prediction of IT jobs with high accuracy.

S. Ball, Tarımer [5] explored an application of ANN for prediction and comparison with statistical methods called the Box-Jenkins Method (BJM) and Regression Analysis (RAM). The dataset of the unemployment rate during 1992-2002(Turkey) is analyzed in connection with the people's educational situations to predict 2002-2012 unemployment rates. Their experiments showed that ANN is quite robust since it can recognize the complex relations from among great number of variables better than other conventional statistical methods: BJM and RAM.

According to the investigation, not only every country needs their respective job prediction model for their national development but also Myanmar needs to build job prediction model to provide the requirements of Myanmar industry. The main objectives of this paper are to build the prediction model and to predict the job category based on Myanmar Labor data. ANN and back-propagation learning algorithm are applied to achieve the best prediction result. For the experimental results, four different models are tested with different activation functions and compared their accuracy.

III. ARTIFICIAL NEURAL NETWORK

For predictive modeling, there have many machine learning methods and techniques. Among them, Artificial Neural Network (ANN) is the most widely used network in several applications. ANN is a neural network that has a high ability to learn and play a great role in solving complex nonlinear problems. This algorithm is closely related to the biological neuron system. ANN, a computational model comprises a huge collection of bridged artificial neurons. To represent relations between inputs and outputs, it is possible to train neurons and their connections with data. One key feature of ANN is its advantages in predicting output or result in the absence of real information knowledge, regarding the modeled system. Computer science, finance, engineering, and many other fields depend on ANN to solve vivid crucial problems. The main reason for deploying ANN in real-time applications is its computational efficiency [1].

The architecture of ANN consists of three main layers called the input layer, hidden layer, and output layer as displayed in Fig 1. First, input layer contains input neurons x_1, x_2, x_3, . . . , x_n. Second, the hidden layer includes hidden neurons called processing elements. Last, the output layer has output neurons and use an activation function to map for the required output form. Activation functions (Transfer functions) are used in hidden layers and output layers except for the input layer. The output of input layer is obtained from the following propagation function and activation function.

$$f(x) = \sum_{j=1}^{n} x_j w_j \quad (1)$$

where x is input, n is number of inputs, w is weight and j is number of neurons.

In neural networks, there have some popular activations called Sigmoid, Tanh, Relu (Rectified Linear Unit), Softmax. The most common activation functions Sigmoid and Tanh are used in the hidden layers as following equation 2,3. The Softmax function is one of the most important output function used in deep learning within the neural networks [7] and can be applied where there are three or more classes of outputs.

- **Softmax** - Combination of multiple sigmoids.
- **Sigmoid** - Transforms the values between range 0 and 1.

$$S(x) = \frac{1}{1+e^{-x}} \quad (2)$$

- **Tanh** - Transforms the values between range -1 and 1.

$$T(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \quad (3)$$

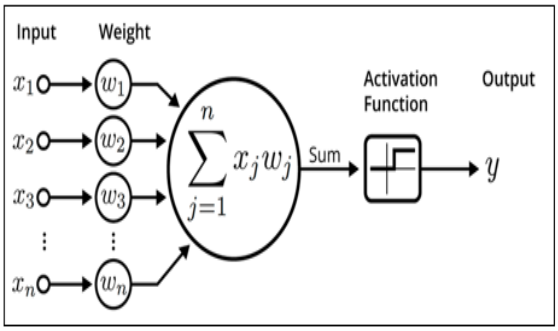

Fig. 1. Components of ANN

IV. PROPOSED SYSTEM

The proposed jobs prediction model is implemented by ANN on real dataset, Myanmar Labor data which is requested from the Department of Labor, Myanmar. The implementation of the proposed model is composed with the following steps and shown in Fig 2**.**

(1) **Data Preprocessing**: Fill in missing values, identify and remove inconsistent features.
(2) **Data Division**: Divide data set into two, training and testing sets.
(3) **Model Building**: Develop Job prediction model using ANN.
(4) **Evaluation Measures**: Accuracy, precision and recall are used to measure the predicted results based on train and test dataset.

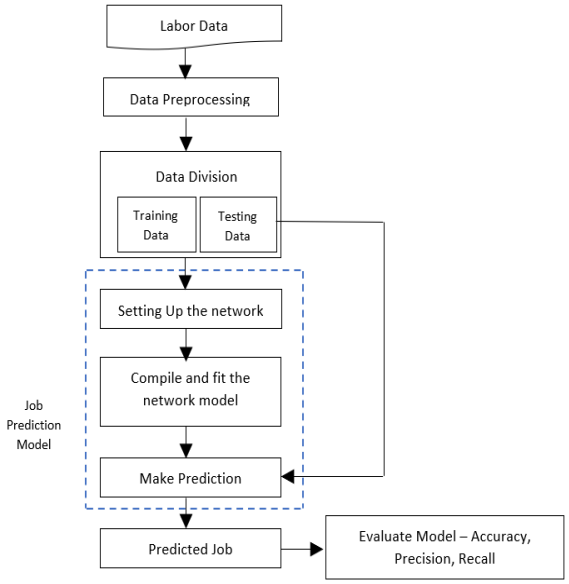

Fig. 2. Proposed system for job prediction

A. Dataset description

The data set used for this study has 12 attributes and over 10000 instances comprising of Myanmar Labor data which are recorded from 2018 to 2019 by the Department of Labor. The attributes used in the experiments and the description of the dataset is presented in Table I.

TABLE I. DATA SET DESCRIPTION

No.	**Attribute Name**	**Data Type**	**Attribute Value**
1.	Urban/Rural	Numeric	1= Urban, 2= Rural
2.	Sex	Numeric	1=Male, 2=Female
3.	Age	Numeric	1 to 99
4.	Education	Numeric	1=Never attend, 2=Below Primary, 3= Primary, 4=Middle, 5=High, 6=Under Graduate, Diploma, 7=Degree and above
5.	Ownership	Numeric	1=Government/Public, 2=Cooperative, 3=Joint-venture (National & Foreign), 4=Joint-venture (Public+Foreign), 5=Joint-venture (public+national), 6=Private (National), 7=Private (Foreign), 8=Private Household, 9=Others
6.	Registration Status	Numeric	1=Ministryof Planning, 2=Ministry of Commerce, 3=Ministry of Hotle and Tourism, 4=Ministry of Health, 5=City Development Committee, 6=Others, 7=No registration required, 8=In the process of registration, 10=Not register, 99=Don't know
7.	Job Experiences	Numeric	1=Less than 3 months, 2= 3-6 months, 3=6-12 months, 4=1-2 years, 5= 2-5 years, 6= 5-10 years, 7=10 years or more
8.	Employment Status	Numeric	1=Parment/regular employee, 2=Casual Employee, 3=Paid apprentice, 4=Employer, 5=Own-account worker, 6=Unpaid family worker
9.	Working days per month	Numeric	From 1 to 30
10.	Working hours per week	Numeric	From 3 to 105
11.	Monthly Wages	Numeric	From 5 to 9000(in thousand)
12.	Occupation Code	Numeric	1,2,3,4,5,6,7,8,9,10

B. Data Preprocessing

The original dataset has many missing and noisy data. To clean the dataset, the duplicated data and missing values are identified and then do the data cleaning. For example, Monthly Wage attribute is empty in some instances, so Working Dpm and Working Hpw attributes of this instance are used to fill empty Monthly Wage attribute compared with other instances.

Data preprocessing is the initial phase of the method. It is intended to transform the raw data to a format that is easier and more effective to use for future processing steps. In the early phase, 11 input features are normalized using Min-Max scaler. Normalization can improve the training time because all data used in the training have the same scale, for example, in the range of 0 and 1. For this purpose, Equation (4) is implemented, where X_{norm} is the result of normalization, X is the initial value before being normalized. Here, X_{max} and X_{min} represent the maximum and the minimum values of each feature, respectively.

$$Xnorm = \frac{X - Xmin}{Xmax - Xmin} \quad (4)$$

C. Data Division

The data used in Artificial Neural Network (ANN) including the back-propagation algorithm is divided into training and testing data according to 80% and 20% respectively.

D. Model Building

There are two steps in job prediction model, setting up the model and compiling and fitting the network model. The job prediction model is built by setting the various parameters for the best prediction model. The experiment settings are shown in Table II.

The 11 input neurons are Urban/Rural, Sex, Age, Education, Ownership, Registration Status, Job Experiences, Employment Status, Working days per month, Working hours per week, Monthly Wages. The 10 output neurons are Managers, Professionals, Technicians and associate professionals, Clerical support workers, Services and Sales Workers, Skilled agricultural, forestry & fishery workers, Craft and related trades workers, Plant and machine operators and assemblers, Elementary occupations, Armed forces occupations as presented in Table I.

In all experiments, there are 11 neurons in input layer and 10 neurons in output layer, weights are randomly generated, the batch size is 64 and epoch is 300. The number of neurons in first, second and third hidden layers are same number. Eg. 4-4,5-5,4-4-4 etc. The proposed model is evaluated by using the Adam Optimization function and Categorical Cross Entropy function as loss function. The example network structures are also shown in following Table II.

TABLE II. ANN PARAMETER SETTING FOR MODEL BUILDING

No.	Network's structure (input- hidden-hidden-hidden output)	Activation function (hidden layers, output layer)
1	11-4-4-10	Tanh, Softmax
	11-5-5-10	
	……	
	11-20-20-10	
	11-25-25-10	
2	11-4-4-10	Sigmoid, Softmax
	11-5-5-10	
	……	
	11-24-24-10	
	11-25-25-10	
3	11-4-4-4-10	Tanh, Softmax
	11-5-5-5-10	
	……	
	11-24-24-24-10	
	11-25-25-25-10	
4	11-4-4-4-10	Sigmoid, Softmax
	11-5-5-5-10	
	……	
	11-24-24-24-10	
	11-25-25-25-10	

V. PERFORMANCE EVALUATION OF JOB PREDICTION MODEL

For evaluation measurement of proposed model, the confusion matrix is applied to look for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. In the case of job prediction, the output target class comprises 10 classes of job. The value of accuracy, precision, and recall for each class are calculated as follows.

Accuracy: percentage between the total number of correctly classified entities (TP + TN) and the whole entities (TP + TN + FP + FN).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Precision: a measure of result relevancy calculates by the proportion of correctly classified entities (TP) among all entities classified (TP + FP).

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall: known as sensitivity to measure how many truly relevant results are returned, which can calculate by the proportion of correctly classified entities (TP) among all entities (TP + FN).

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

A. Experimental Results

The combination and tuning of ANN parameters can achieve the best predicted result. The data were divided into training and testing sets as 80% and 20% respectively.

The data used in the multilayer feed-forward neural network (ANN) uses a back-propagation algorithm that is arranged in a format of 11 input parameters and training for 3643 instances and testing for 911 instances. The output of the proposed system is one of ten job classes.

This paper selects the best model among four different models with different parameters. The experimental results are shown with accuracy, micro-

average precision, and micro-average recall as following Fig 3 to Fig 6.

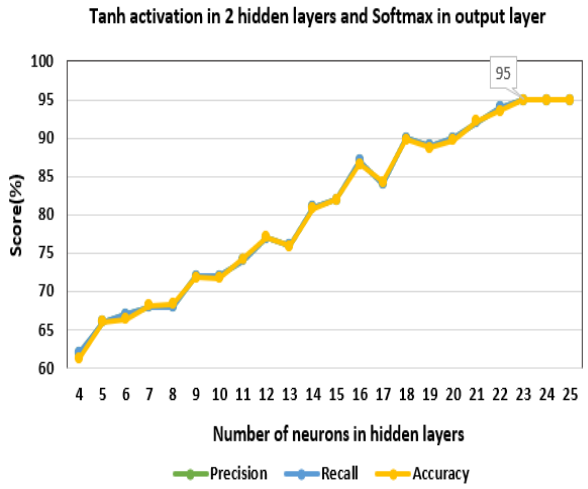

Fig. 3. Proposed system's result for Experiment 1

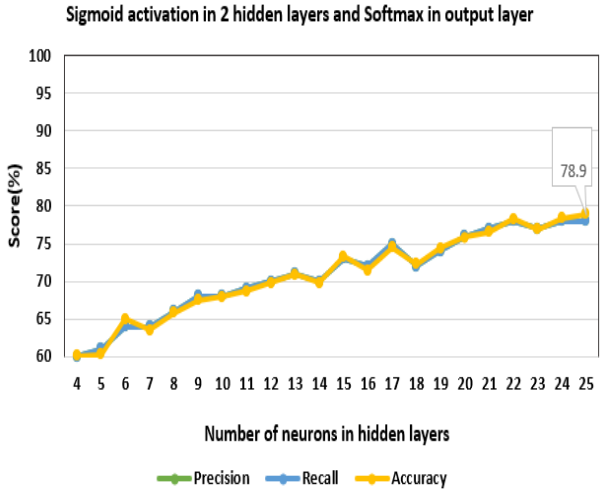

Fig. 4. Proposed system's result for Experiment 2

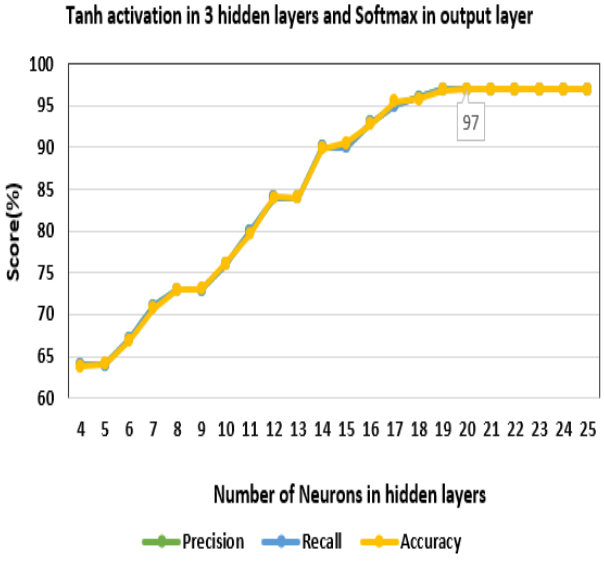

Fig. 5. Proposed system's result for Experiment 3

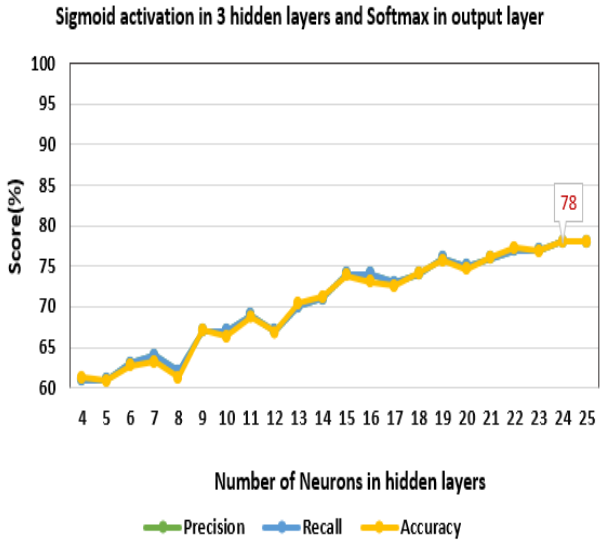

Fig. 6. Proposed system's result for Experiment 4

According to the observation and the classification results on each label, Experiment 3 achieved the best result and is presented in Table III. In this model, the class "Services and sales workers", "Skilled agricultural, forestry & fishery workers", "Craft and related trades workers" achieved the highest Precision and Recall and the class "Armed forces occupations" got the lowest Precision 88% and Recall 50%.

TABLE III. EXPERIMENTAL RESULTS OF PRECISION, RECALL AND ACCURACY FOR EACH OUTPUT LABEL

No.	Job Category	Precision (%)	Recall (%)	Accuracy (%)
1.	Managers	96	100	97
2.	Professionals	94	91	
3.	Technicians and associate professionals	80	100	
4.	Clerical and support workers	95	96	
5.	Services and sales workers	100	100	
6.	Skilled agricultural, forestry & fishery workers	100	100	
7.	Craft and related trades workers	100	100	
8.	Plant and machine operators and assemblers	100	83	
9.	Elementary occupations	98	97	
10.	Armed forces occupations	88	50	

When the predictability of the proposed model is examined, increasing the number of neurons in hidden layers gets the higher the prediction accuracy. The four selected models from all experiments are shown which obtained the highest accuracy scores. They show the best accuracy when the number of neurons is up to 20 and Tanh activation function in all hidden layers. The

accuracy is not apparently increased when the Sigmoid activation function is used in all hidden layers. In contrast, when increasing the number of neurons above 20 and Sigmoid in all hidden layers, the accuracy is not immediately increased and still stable, indicating that the number of neurons and Tanh activation affects the predictive results. The experimental results are shown in Fig 3 to 6.

As the above experiments, the highest accuracy score on each output label can be seen in Fig 5. The network structure is one input layer, three hidden layers, and one output layer which is 11 input neurons in the input layer, 20 hidden neurons in the first hidden layer, 20 hidden neurons in the second hidden layer, 20 hidden neurons in the third hidden layer, 10 output neurons and Tanh activation in hidden layers.

VI. CONCLUSION AND FUTURE WORK

In Myanmar, the job prediction model is an essential thing to help both employees and employers of every organization. Many experiments are done to get the best performance before selecting the best model. This paper presents the experimental results of four different models which are built with two and three hidden layers with Tanh or Sigmoid activation, and Softmax in the output layer. They are the top performance results of the experiments. Among them, the network's structure of 1-20-20-20-10 is the best model with an accuracy of 97%. When the combinations of ANN parameters are identified and developed in the model, the experimental results show that the proposed prediction model in Experiment 3 has the best prediction effect.

Although the proposed system is trained on a real dataset of Myanmar Labor, it will have some limitations because it depends on the available data. Therefore, the accuracy of the proposed system will improve and can predict more job categories identified by ISCO when more efficient data are obtained. Job prediction will help the significant improvements of experience and skills of job seekers, and employers to bring more transparency and to discover the market qualification of different job positions.

ACKNOWLEDGEMENT

I would like to thank Dr. Saw Sanda Aye, Rector of University of Information Technology for her kindly support to request Myanmar labor's data from the Department of Labor. I also sincerely thank the Department of Labor, Ministry of Labor and Immigration located in Nay Pyi Taw, who provided me the required data for my thesis. I would also like to extend my thanks to the members of the Data Analysis and Management laboratory.

And most importantly I would like to thank my parents for supporting me and believing in me in all the difficult moments.

REFERENCES

[1] Akku George Saju, "Future Job Prediction in Trivandrum using Machine Learning Techniques", 2018.

[2] Anifat Olawoyin, Yangjuin Chen, "Predicting the future with Artificial Neural Network", Complex Adaptive Systems Conference with Theme: Cyber Physical Systems and Deep Learning, pp. 383–392, 2018.

[3] Kuo-Yan Wang, Hsien-Yu Shun, "Applying Back Propagation Neural Networks in the Prediction of Management Associate Work Retention for Small and Medium Enterprises", Universal Journal of Management, pp. 223-227, 2016.

[4] Nedhal A. Al-Saiyd, Amjad S. Al-Takrouri, "Prediction of IT Jobs Using Neural Network Techniques", Ubiquitous Computing and Communication Journal, 2015.

[5] S. Ball, Tarımer, "An Application of Artificial Neural Networks for Prediction and Comparison with Statistical Methods", Elektronika Ir Elektrotechnika, VOL. 19, pp. 1392-1215, 2013.

[6] International Standard Classification of Occupations (ISCO), <https://www.mol.gov.mm/mm/information/statistics/>.

[7] https://medium.com/data-science-bootcamp/softmax-function-beyond-the-basics-51f09ce11154