M.Sc. Engg. Thesis

# Detecting Geo-Financial Solvency Using Social and Web Media Networks

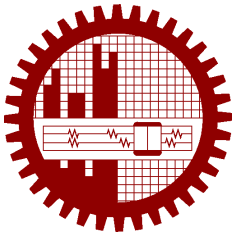

Submitted by

Md. Aflatul Sagir

0413052027

Supervised by

Dr. Mahmuda Naznin


Submitted to

**Department of Computer Science and Engineering**

**Bangladesh University of Engineering and Technology**

Dhaka, Bangladesh


in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

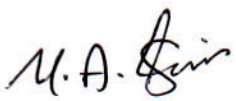
September 2019

*I dedicate this thesis to my late parents.....*

# Candidate's Declaration

I, do, hereby, certify that the work presented in this thesis, titled, **"Detecting Geo-Financial Solvency Using Social and Web Media Networks"**, is the outcome of the investigation and research carried out by me under the supervision of Dr. Mahmuda Naznin, Professor, Department of CSE, BUET.

I also declare that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.



Md. Aflatul Sagir

0413052027

The thesis titled "**DETECTING GEO-FINANCIAL SOLVENCY USING SOCIAL AND WEB MEDIA NETWORKS**", submitted by Md. Aflatul Sagir, Student ID 0413052027, Session April 2013, to the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, has been accepted as satisfactory in partial fulfilment of the requirements for the degree of Master of Science in Computer Science and Engineering and approved as to its style and contents on September 28, 2019.

## Board of Examiners

1. Dr. Mahmuda Naznin — Chairman (Supervisor)
   Professor
   Department of CSE, BUET, Dhaka

2. Dr. Md. Mostofa Akbar — Member (Ex-Officio)
   Professor and Head
   Department of CSE, BUET, Dhaka

3. Dr. Abu Sayed Md. Latiful Hoque — Member
   Professor
   Department of CSE, BUET, Dhaka

4. Dr. A. B. M. Alim Al Islam — Member
   Associate Professor
   Department of CSE, BUET, Dhaka

5. Dr. Hasan Sarwar — Member (External)
   Professor
   Department of CSE
   United International University (UIU), Dhaka

# Acknowledgement

First of all, I would like to express my heart-felt gratitude to my supervisor, Prof. Dr. Mahmuda Naznin, for her constant supervision of this work. She helped me a lot in shaping, deciding steps of this work, and in providing infrastructural supports.

The door to Prof. Dr. Mahmuda Naznin office was always open whenever I ran into a trouble or I had a question about my research or writing. She consistently steered me in the right the direction whenever she thought I needed it.

I would like to thank the honorable members of my thesis committee: Prof. Dr. Md. Mostofa Akbar, Prof. Dr. Abu Sayed Latiful Hoque, Prof. Dr. Hasan Sarwar and Dr. A.B.M Alim Al Islam for their encouragement, insightful comments, and valuable suggestions.

Finally, I must express my very profound gratitude to my family members and friends for providing me support and continuous encouragement throughout my years of study and through the process of completion of the research. This accomplishment would not have been possible without them.

Dhaka
September 28, 2019

Md. Aflatul Sagir
0413052027

# Contents

# List of Figures

# List of Tables

# List of Algorithms

# Abstract

Financial solvency analysis of a geographic area is a very challenging task because it costs lots of money and a big volume of data handling is involved. Every country in the world checks its financial status, which is known as the economic census. There are some techniques to measure financial status like using of some questionnaires where data is collected from field level manually, or using financial statements. These techniques are old and these use data collection manually which increase the cost, money and effort. The use of social and web media network has been increased a lot. As a part of socio-activities people share their personal information like occupation, home town, events, plan with his/her friends on social media. On the other hand, web media or online access of information provide information for economic establishments (e.g. public institutions, private companies, etc). There are many research work like cultural boundary detection, community detection, food consumption statistics, sentiment analysis, and personality predicting, have been done using social media. It is very important to have economic census for a country. In this study, an automated economic zone finding framework has been designed using social-sensing and web mining. Here, data has been collected using social and web media. By using the proposed framework, the financial status of eight major divisional cities of Bangladesh has been successfully detected. And the last of all, the city's ranks have been measured according to their financial solvency. This technique is automated, so it requires less time and it costs less money than other traditional surveyed methods. According to best of our knowledge, this is the first technique in our country that uses the web and social media to verify the financial status of different cities and this technique performs better than existing surveyed approaches.

# Abbreviations

| | |
|---|---|
| **AHE** | *Activities of households as employers* |
| **AER** | *Arts, entertainment and recreation* |
| **AEOB** | *Activities of extraterritorial organizations and bodies* |
| **AFF** | *Agriculture, forestry and fishing* |
| **AFSA** | *Accommodation and food service activities* |
| **ASSA** | *Administrative and support service activities* |
| **BBS** | *Bangladesh Bureau of Statistics* |
| **Bus.** | *Business* |
| **BP** | *Buying or Purchasing* |
| **Const.** | *Construction* |
| **DFA** | *Daily Financial Activities* |
| **ED** | *Eating and Drinking* |
| **Edu.** | *Education* |
| **EE** | *Economic Establishment* |
| **EGSA** | *Electricity, gas, steam, and air conditioning supply* |
| **Emp.** | *Employment* |
| **Ent.** | *Entertainment* |
| **FIA** | *Financial and insurance activities* |
| **GDP** | *Gross domestic product* |
| **HHSWA** | *Human health and social work activities* |
| **IC** | *Information and Communication* |
| **LR** | *Linear Regression* |
| **MAQ** | *Mining and quarrying* |
| **MFG** | *Manufacturing* |
| **NIPA** | *National income and commodity accounting* |
| **NR** | *Non Residents* |
| **Occ.** | *Occupation* |
| **OSA** | *Other service activities* |
| **OO** | *Other Occupation* |

**PAD** *Public administration and defense; compulsory social security*
**PPI** *Producer price index*
**PSTA** *Professional, scientific and technical activities*
**REA** *Real estate activities*
**SMO** *Sequential minimal optimization*
**SSM** *Solvency Sensing Method*
**SVM** *Support Vector Machine*
**TAS** *Transportation and Storage*
**TNU** *Total Number of Users*
**TNP** *Total Number of Posts*
**TNNR** *Total Number of Non-Residents*
**VP** *Voted Perceptron*
**WRT** *Wholesale and retail trade; repair of motor vehicles and motorcycles*
**WSWRA** *Water supply; sewerage, waste management and remediation activities*

# Chapter 1

# Introduction

*Geo-Financial Solvency* measurement verifies the financial status of the people residing in a particular geographic area. Due to its importance, every country in the world performs checking of its financial status of the country's different locations in a certain period. It is an important activity with the aim of improving the quality of life based on economic statistics. Businesses use *Economic Census* data to help to make decisions about where to expand the business, where to set up a new business, how much to produce etc. To compare the performance to other businesses in the industry or community *Economic Census* plays a vital role. Local communities use *Economic Census*'s results to attract new businesses, assess the economic health of their localities, to understand the characteristics of their business base, and to compare their community to other geographical areas. Individuals can use Economic Census results to identify emerging job markets and growing industries.

## 1.1 Motivation

Different types of tools and techniques are used to measure the financial health status of any geographic location. One of the most common techniques is to use of some questionnaire [1, 2], about the type of economic activities of the business units, type of the manufactured products or services, type of ownership, equity, sales and accounting practices. To accomplish such a process, information is collected from the field level (end users) manually.

In case of an organization, the financial solvency is verified from the balance sheet received from the financial statements of that organization [3,4]. In that case, the financial solvency of the institution is determined using some accounting theories [5, 6], based on the total assets and total liabilities of that organization. Those techniques cost huge money and require huge time.

Some international organizations [7–10] use their own formulas and tools to measure the

financial conditions of different countries. Very recently Support Vector Machine (SVM) technique has been used to analyze data got from the third party [11]. These types of techniques reduce processing time after data is collected. However, the total time for data collection is not reduced.

Financial solvency measurement is a challenging task because:

- It costs lots of money.
- A big volume of data handling is involved.
- A large volume of man-power is required.

This can be handled if the following steps are done.

- Data collection from social and web media can reduced cost of money.
- Use of an automated tools can save processing time and a few people to operate the software.

To minimize the financial cost of conducting a census, an automated framework has been proposed in this thesis, where data has been collected from social and web media. To reduce data processing time, machine learning has been used in this framework.


Figure 1.1: The process of automatic geo-financial solvency detection using SSM.

The figure 1.1 shows the process of geo-financial solvency detection using proposed method what we call *Social Sensing Machine* (*SSM*). In the first step, the data is downloaded from social and web media. In the second step, the downloaded data is entered into SSM for analysis. The last part of this figure compares the score of the targeted areas and ranks them according to the score value.

## 1.2 Challenges

In this section, we talk about some challenges in developing automated solvency framework. These are stated as follows:

1. *Data Extraction:* The social media data is too noisy. Therefore, the financial data extraction from numerous non-related data in social media can be tricky.
2. *Structured Data:* Making unstructured data into structured format is one of most difficult task in data processing. The success of this method depends on the structured data.
3. *Indicator Selection:* Choosing the right indicators for the dimensions and the proper weight distribution among the indicators is one of the most difficult tasks. In this case, we followed the international methods.
4. *Computation Method:* Finding a suitable score calculation method that will reduce the computation time in the total working process.

## 1.3 Applications and Advantages

There are many applications and advantages of this technique. Some applications of this method are pointed below:

1. **Economic Census Measurement:** The proposed framework can be used for area wise economic census measurement in national or international level with a very low cost and less time.
2. **Financial Status Measurement:** The proposed technique can be used to measure person's financial status by analysing daily financial activities through social media.
3. **Public Opinion Verification:** To verify the public opinion to start a new business in a city or in a country, this method can be helpful.
4. **Measurement of Product Demand:** Using SSM, the public demand of certain types of products can be measured.
5. **Dominant Factor Finding:** This method can help to find out the most dominant factors that makes an area most or less solvent. (e.g., manufacturing, non-residents etc.)

There are many advantages in the proposed method compare to traditional surveyed methods, some of them are describes in Table 1.1.

| Criteria | Traditional Method | SSM |
|---|---|---|
| Cost | Huge money investment needed | Less money investment needed |
| Processing time | Need more data processing time | Need less data processing time |
| People | More people needed to engaged | As it fully automated software based, need a few people to operate. |

Table 1.1: Advantages of the proposed method compared to conventional survey methods

## 1.4 Scope of The Work

This research has been designed for eight divisional cities of Bangladesh (1) Barishal (2) Chattogram, (3) Dhaka, (4) Khulna, (5) Mymensingh, (6) Rajshahi, (7) Rangpur, and (8) Sylhet.
The economic establishment data has been collected by using three websites and a search engine such as (1) https://www.bangladeshyellowpages.com, (2) https://bangladesh.gov.bd, (3) https://en.wikipedia.org, (4) https://www.google.com. This research has covered all economic units- permanent and temporary establishments, household based economic activities which are available at web media. It should be noted that some establishments engaged illegal activities like bar, smuggling, gambling, beggary, prostitution, etc. are kept out of this research.
This research uses social media data set which has collected from Facebook. In case of selecting user posts, we ignore the posts which are not written in English and the posts which are irrelevant to our research topic.

## 1.5 Contributions

The major contributions in this thesis are stated as follows.

1. *Area Ranking:* To the best of our knowledge, we have ranked eight divisional cities of Bangladesh according to their financial status using social and web media data.

2. *Detecting Dominant Factors:* We have been able to figure out the most influential factors those affect a region's financial stability. For example, the *"Non-Resident"* indicator enables the Sylhet city to gain better rank in the occupation and *DFA* described later.

3. *Financial Data Extraction:* Financial activities have been extracted from numerous non-related plain texts of a post.

4. **New Dimension Addition:** To verify financial solvency using social and web media data, new financial impact creating features have been created.

5. **Use of Modern Tools:** For data processing and mining, *R*, machine learning tools *WEKA* have been used for geo-financial analysis.

6. *Dictionary Construction:* Four dictionaries which contain 10000 words for solvency analysis have been built.

7. **Auto Framework Build-up:** A solvency sensing framework has been built up which can automatically detect geo-financial solvency.

## 1.6 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 discusses the existing work related to geo-financial survey methods and challenges. It highlights the scope of the works based on the limitation found in the state of the art methods. This chapter discusses benchmark solutions to measure financial stability by some standard organization or other countries. Chapter 3 explains the detail framework of our proposed method called *Solvency Sensing Method (SSM)*. The assumptions of this research is also discussed here. The mathematical models used for the score and rank calculation are explained at the end of this chapter. Results and analysis are discussed in Chapter 4. The tools and techniques, data-set are given in this chapter. The comparative result analysis and limitations of this research are also discussed here. The summary of the results are shown in the end of this chapter. Finally, in Chapter 5, we conclude the thesis by summarizing the proposed work and highlighting some future scopes.

# Chapter 2

# Related Work

There are many solutions which have been proposed to measure financial stability of an organization or a country. We divide them into three categories based on - (1) Financial Theorem (2) Census, and (3) Organization's Own Formula. Here, we also discuss some work done using social media network.

## 2.1 Financial Theorem Based Methodology

The financial theorem based method uses different financial theorems for analysis organizational financial stability. It uses different types of dimension and indicators. The most common types of dimension: (1) Cash solvency (2) Budgetary Solvency (3) Long run Solvency (4) Service-level solvency.
(1) *Cash solvency* [3] describes a government's (or organization's) ability to generate sufficient financial resources to pay it's current liabilities. The working capital to expenses ratio is a indicator of cash solvency that measures the liquidity. The corresponding equation is as follows:

$$\text{(i) The Working Capital to Expenses Ratio} = \frac{\text{Current Assets - Current Liabilities}}{\text{Expenses}}$$

(2) *Budgetary solvency* [4] demonstrates a government's financial ability to maintain current or desired service levels within a period by sufficiently funding operating expenses. There are three indicators to measure Budgetary solvency: (1). the continuing services ratio, (2). the fund balance ratio, and (3). the operating surplus ratio. *Continuing services ratio* is an indicator that examines unrestricted net assets as a percentage of expenses for all funds government-wide.

$$\text{(ii) The Continuing Services Ratio} = \frac{\text{Unrestricted Net Assets}}{\text{Total Expenses}}$$

The *fund balance ratio* is an indicator that compares unrestricted general fund balance to general fund expenditures, reflecting the operating savings that a government has accumulated relative to the government's operating expenditures for that fiscal year.

$$\text{(iii) Fund Balance Ratio} = \frac{\text{Unrestricted General Fund Fund Balance}}{\text{General Fund Expenditures}}$$

The *operating surplus (deficit) ratio* shows the general fund operating surplus or deficit as a percentage of total operating expenses. This indicator reflects the difference between revenues and expenditures in completed fiscal years. Generally, a larger ratio reflects excess operating revenues over expenditures. Therefore, a ratio greater than zero and an increasing trend are considered favorable.

$$\text{(iv) Operating deficit ratio} = \frac{\text{General Fund Surplus or Deficit}}{\text{Net Operating Expenditures}}$$

(3) *Long-run solvency* [5] assesses the availability of future resources to pay for existing long-term obligations. The *net worth ratio* is an indicator that measures restricted and unrestricted net assets as a percentage of total asset. It is a measure of the net worth of a government and signifies the government's ability to pay off existing long-term liabilities. A larger ratio indicates more accessible resources for the government and therefore, an increasing trend and higher ratio are considered favorable.

$$\text{(v) Net Worth Ratio} = \frac{\text{Restricted and Unrestricted Net Assets}}{\text{Total Assets}}$$

The *debt service expenditure ratio* is the percentage of debt service expenditures out of total governmental fund expenditures. The indicator can be used to evaluate service flexibility by determining the amount of expenditure committed to the annual debt service. With a higher debt service expenditure ratio, a larger portion of expenditures is being allocated to paying for debt issued by the government rather than regular government services. A lower ratio and decreasing trend are considered favorable.

$$\text{(vi) Debt Service Expenditure Ratio} = \frac{\text{Debt Service Expenditure}}{\text{Total Expenditures}}$$

(4) *Service-level solvency* [6] which reflects a government's ability to maintain services at the quality and level required to ensure the safety and welfare of citizens and to meet their expectations and desires. Expenses per capita divide the total expenses of the primary government, which include governmental activities and business-type activities, by population. Higher expenses per capita reveal a more expensive government and lower solvency to sustain that expense level. Therefore, a lower ratio than the average of all of the cities and decreasing trend are considered favorable.

$$\text{(vii) Expenses per Capita} = \frac{\text{Total Primary Government Expenses}}{\text{Population}}$$

*Liabilities per capita* divides the total liabilities of the primary government by population and represents the government's relative indebtedness with regard to future taxpayers. A lower ratio and decreasing trend are considered favourable. Since higher than average or increasing liabilities can be a cause for concern, a lower ratio than the average of all of the cities and decreasing trend are considered favorable.

Taxes and fees per capita divides all taxes and charges for services for primary government activities by population. Higher taxes and fees per capita reflect a higher tax burden for residents and a lesser ability of the government to raise taxes or fees further to sustain current service levels. As such, a lower than average ratio and decreasing trend are considered favorable.

$$\text{(ix) Taxes and Fees per Capita} = \frac{\text{Total Primary Government Taxes and Charges for Services}}{\text{Population}}$$

## 2.2 Economic Census Based Methodology

There are many approaches are used to measure economic census. We describe some of them. Bangladesh Bureau of Statistics (BBS) conducted economic census in 2013 [1] to measure financial condition of Bangladesh by calculating the number of economic households, establishments and employee's work on those organizations. They generate some questionnaire about economic establishments, households and employees. The data was collected manually by filling up the form. These census costs more than 500 crore and total processing and outcome time was more than 1 year.

Central Statistics Office (CSO) [2], conducted economic census of India 2013, they includes various agricultural and non agricultural activities, they divides all economic activities into 23 categories, outcome of that census was (1) total number of establishments counted is about 58.5 million, (2) 59.9% of the establishments belong to rural areas, (3) 20.5%

of the establishments operate from outside household without fixed structure, (4) total number of persons employed is about 128 million with rural share of 51.9%, (5) percentage of female workers is 25.6%, (6) growth rate in number of establishments over fifth EC (2005) is 41.7%, (7) growth rate in total employment over fifth EC (2005) is 34.4%. It was also very costly and time consuming.

## 2.3 Organization's Own Formula Based Methodology

The international organizations like World Bank, UNDP, World Economic Forum uses their own formulas and tools to measure the financial condition of different countries in the world [7–10].

The Civic Federation [3, 5], compared the relative fiscal performance of the city of Chicago to 12 other major U.S cities using nine financial indicators compiled from audited financial statements from FY2007 through FY2011. Primarily, rank was counted for each indicator based on five year financial change of the cities. The final rank is counted as the average of the indicators' ranks. The city which has the lowest total average indicator rank marked as Rank 1 and so on. In their report, Columbus (Ohio) scored Rank 1, Chicago ranked 11 and Detroit ranked 13.

The World Economic Forum [8], analysis the economic development of 140 countries in the world. They divided the development into three stages. (1) Factor driven (35 countries), (2) Efficiency driven (31 economies) and (3) Innovation driven (38 economies). *Stages* were calculated by three sub indexes: Basic requirements, Efficiency enhancement and Sophistication factors. There is no uniform and universal rule for measuring economic census. Most of the cases, data is collected manually which is costly and time consuming. It should be noted that, the results depend on the method what is adopted.

## 2.4 Social Media

Silva et al. [12] describes a technique of identifying people's cultural boundaries and similarities at different scales based on analysis of Foursquare check-ins. Firstly, they mapped food and drink related check-ins that is extracted from Foursquare into user's cultural preferences. Secondly, they identified specific personal preferences, such as the taste of a certain type of food or drink, e.g., pizza, coffee, as well as temporal habits, such as the time and day of the week when an individual visiting a restaurant. Thirdly, they analyzed this data to measure the cultural distance between two countries, cities or regions. Then they applied a simple clustering technique (k-means), using this cultural

distance measure, to draw cultural boundaries across countries, cities or regions.

Yang et al. [13] developed a model called CESNA (Communities from Edge Structure and Node Attributes) for detecting overlapping communities in networks with node attributes. In which *Adjacency Matrix* and values of nodes in a graph network were generated from *Node Community Memberships*. Then by adopting a *block coordinate ascent approach* communities were identified.

Qi et al. [14] presented a technique of community detection which has been formulated by integrating contents of the edges of graph network along with graph structure through *matrix factorization* to derive a high quality latent vector representation for the edges based on an analysis of the link matrix. That representation of the latent vectors is used to extract communities by applying k- means clustering. known as EIMF (Edge Induced Matrix Factorization).

Hoang et al. [15] extracts sentiment from micro event on twitter. They at first found micro events about transportation system of Singapore, then analyzed the data, and found out the sentiment of transport and road communication system development.
Social media data can play an important role to find the financial status of the people of particular geographic locations,.

# Chapter 3

# Solvency Sensing Method

The main goal of this research is to provide geo-financial ranking. As a sample, this ranking has been done for divisional cities of Bangladesh according to the financial status by analyzing social media and web media networks. To achieve this goal, we use three dimensions with twenty eight indicators. Then, the score value of dimensions and indicators has been measured on a scale of 1-100. The city with the highest score earned to get the first position in the ranking. This research has been conducted on the data of eight divisional cities of Bangladesh- Barishal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet.

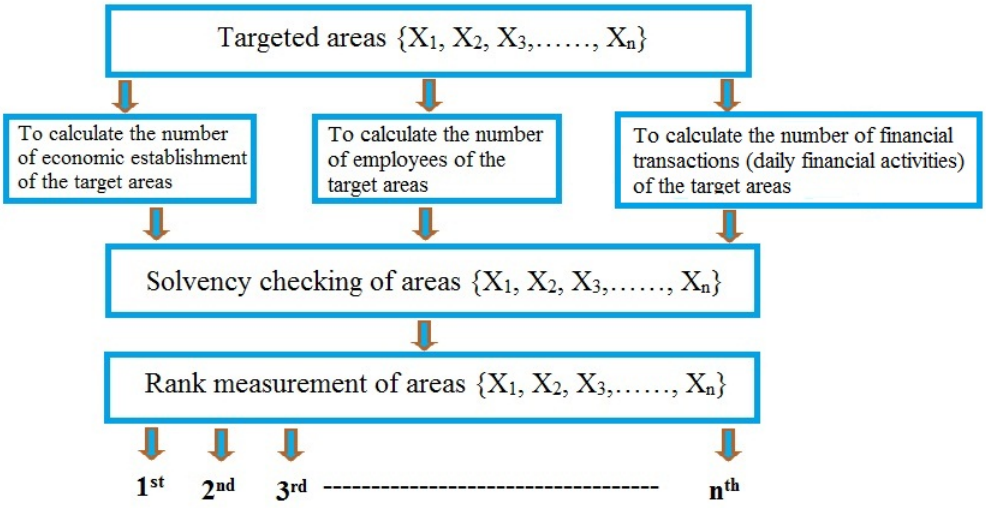



Figure 3.1: Financial Solvency measuring at researcher's view.

The figure 3.1 depicts the number of economic establishments from web media, information on employment, and different financial events from social media. The economic establishments are considered here garments, textile, bank, transport, airlines and any kind of institutions w have money transactions. The number of employees has been

computed from the Facebook user's profile. The events related to financing detected from the user's posts. Scores have been calculated based on the number of establishments, employees, and events in a city. Lastly, the rank has been computed based on the score value.

## 3.1 Assumptions

There are some of our research described as follows.

1. *The area which has more economic establishment is more solvent.* We consider *Economic Establishment* as an asset. We think that the area which has more economic establishments, has more assets.

2. *The area which has more employed persons is more solvent area.* More employed person indicates the geographic area's more economic activities which results into solvency.

3. *The area which has more daily transaction indicates as the area with better financial status.* For example, if the citizens of an area have better purchasing capabilities than other cities, it is considered as more solvent area.

## 3.2 Detail Framework

Figure 3.2 describes the detailed framework of our proposed method Social Sensing Machine or SSM. There are five components in this framework. The first part involves in data collection from the social and web media, the second part involves in data cleaning, the third part involves in preparing the data suitable for score calculation, the fourth part involves in score computation for each dimension and indicator for each city, and the last part computes ranking of the areas according to the achieved score value.

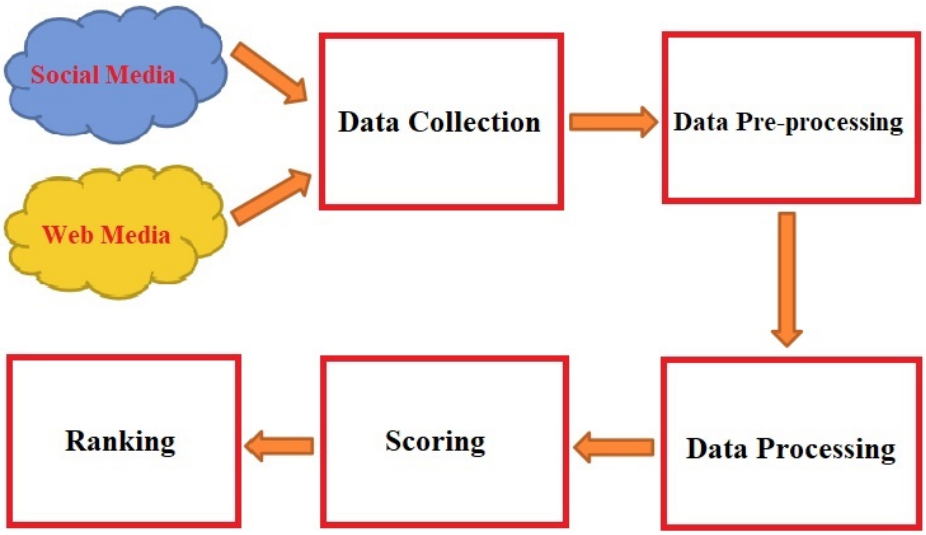



Figure 3.2: Solvency detection framework.

Therefore, our research has been conducted using three economic dimensions and their 28 indicators. The dimensions are as follows:

1. Economical Establishments (*EE*)
2. Occupation (*Occ.*)
3. Daily Financial Activities (*DFA*)

Figure 3.3 describes a sample scenario of area ranking. In the first part, two types of data (Economic Establishment related data and the social media data or user posts) are used as input. In the second part, the data is pre-processed for score calculation. The score is computed over some dimensions based on the indicator data. In the last part, ranked cities have been reported based on the score.

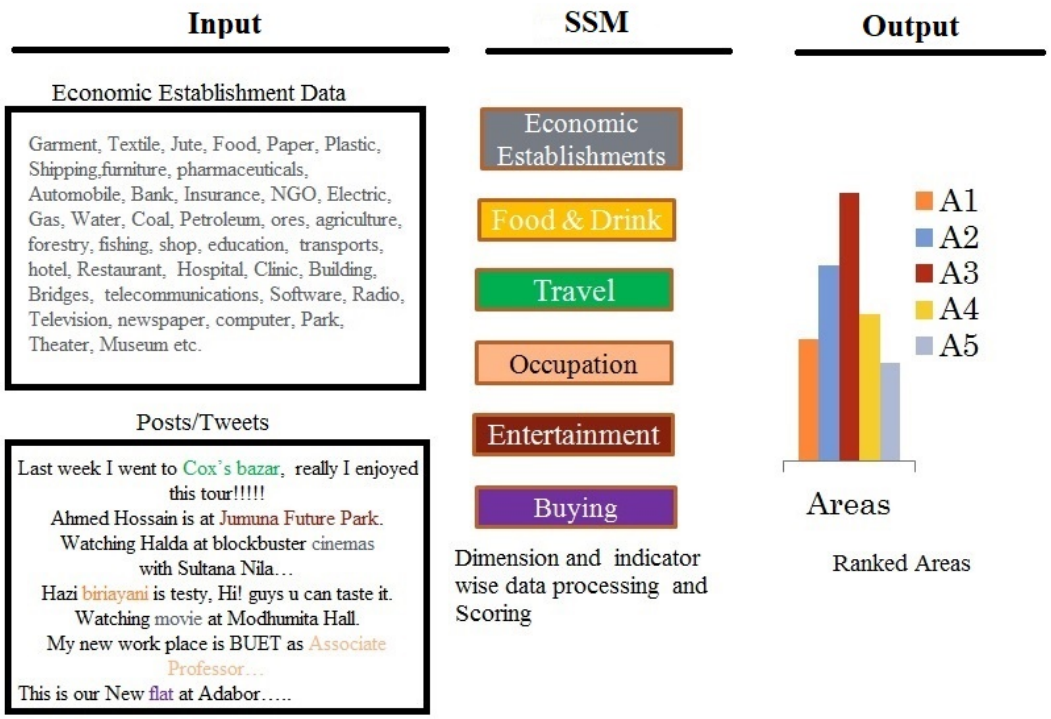



Figure 3.3: A sample scenario of geo-financial ranking using SSM.

## 3.3 Economic Establishment (EE)

An economic unit operates in a permanent or temporary place. Examples: hotel, restaurant, shops, ships, textile mills, banks, insurance etc. According to International Standard (ISIC), all economic activities are classified into 21 broad categories [16]. Those categories have been considered as indicators in this research. The objective is to count the number of economic establishment number of eight cities. The indicators of this dimension are describes in Table 3.1.

| *Dimension* | *Indicators* |
|---|---|
| **Economic Establishment** | 1. Agriculture, forestry and fishing (AFF)<br>2. Mining and quarrying (MAQ)<br>3. Manufacturing (MFG)<br>4. Electricity, gas, steam & air conditioning supply (EGSA)<br>5. Water supply; sewerage, waste management and remediation activities (WSWRA)<br>6. Construction (Const.)<br>7. Wholesale and retail trade (WRT)<br>8. Transportation and storage (TAS)<br>9. Accommodation and food service activities (AFSA)<br>10. Information & Communication (IC)<br>11. Financial and insurance activities (FIA)<br>12. Real estate activities (REA)<br>13. Professional, scientific and technical activities (PSTA)<br>14. Administrative and support service activities (ASSA)<br>15. Public administration and defense; compulsory social security (PAD)<br>16. Education (Edu.)<br>17. Human health and social work activities (HHSWA)<br>18. Arts, entertainment and recreation (AER)<br>19. Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use (AHE)<br>20. Activities of extraterritorial organizations and bodies (AEOB)<br>21. Education (Edu.) |

Table 3.1: Defining indicators of *EE* dimension.

A sample scenario of city ranking at *EE* dimension has been given in Figure 3.4.

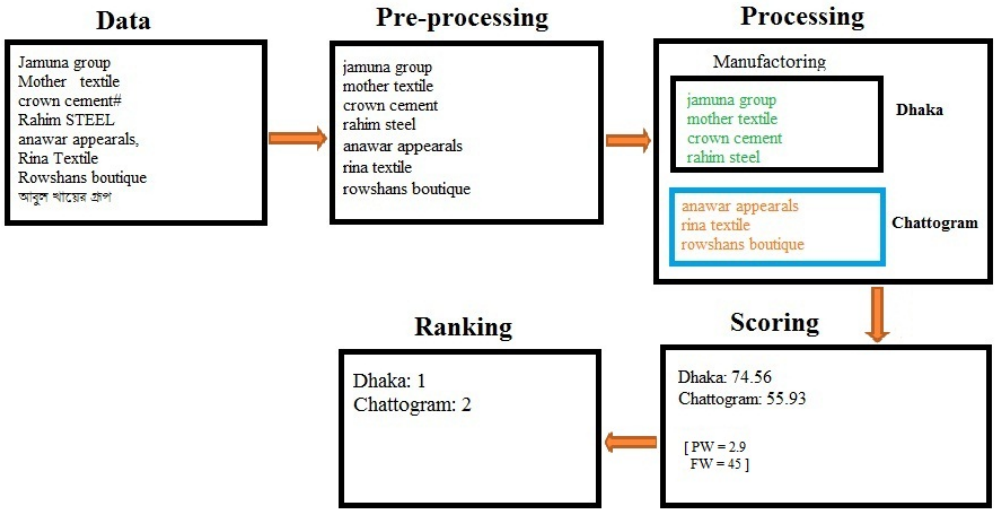



Figure 3.4: A sample scenario of geo-financial ranking at *EE* dimension.

### 3.3.1 Agriculture, forestry and fishing

This indicator includes the activities comprise the natural resources of animal and plants, raising and breeding of animals, comprising the activities of growing of crops, animal and animal product, harvesting of timber and other plants. We collects the firm or company name that performs the activities like:

1. Animal and crop production, and hunting related service activities
2. Forestry
3. Fishing and aquaculture

In the economic census in 2013, the agriculture sector was not included. In this research, we include this important sector. The organizations types have been given in Appendix B.1.

### 3.3.2 Mining and Quarrying

This indicator includes the extraction of minerals occurring naturally as gases (natural gas), liquids (petroleum) or solids (coal and ores). Extraction can be achieved by different methods such as seabed mining, underground or surface mining etc. This indicator also includes supplementary activities for the preparation of crude materials for marketing, for example, grinding, crushing, drying, cleaning, sorting, accumulation of solid fuels, and natural liquid gas.

We collects the firms names that engaged in:

1. Mining of coal and lignite
2. Extraction of crude petroleum and natural gas
3. Mining of metal ores
4. Quarrying of stone, sand and clay
5. Mining support service activities

The organizations types has given in Appendix B.2.

### 3.3.3 Manufacturing

This indicator includes the physical or chemical conversion of materials, substances or ingredients into new products, although it cannot be used as a single universal standard for defining production. Transformed materials, substances or ingredients are raw materials that are used as products in agriculture, forestry, fishing, excavation or mining and other manufacturing activities. Substantial changes, renovations or product remodeling are generally considered as manufacturing. However, there are units that convert materials or substances into new products at home or in the worker's home, and those are selling products to general public. For example, as custom tailors, bakery etc. are also included in this section. We select the companies which are engaged in:

1. Manufacture of food products and beverages
2. Manufacture of tobacco products
3. Manufacture of textiles and wearing apparel
4. Manufacture of leather and wooden products
5. Manufacture of paper and paper products
6. Printing and reproduction of recorded media
7. Manufacture of coke and refined petroleum products
8. Manufacture of chemicals and chemical products
9. Manufacture of medicinal,chemical, pharmaceuticals, and botanical products
10. Manufacture of rubber and plastics products
11. Manufacture of other non-metallic mineral products

12. Manufacture of basic metals and fabricated metal products
13. Manufacture of computer, electronic and optical products
14. Manufacture of electrical equipment
15. Manufacture of machinery and equipment
16. Manufacture of motor vehicles, trailers and semi-trailers
17. Manufacture of other transport equipment
18. Manufacture of furniture
19. Manufacture of jewellery, bijouterie and related articles
20. Manufacture of musical instruments
21. Manufacture of sports goods, games and toys
22. Manufacture of medical and dental instruments
23. Repair and installation of machinery and equipment

According to BBS, the contribution of Manufacturing sector to GDP 18.99% in fiscal year 2017-18. The organizations types details has given in Appendix B.3.

### 3.3.4 Electricity, Gas, Steam, and Air Conditioning

This indicator includes electric power, natural gas, steam, hot water supply activities and through the permanent infrastructure of the line, main and pipe. This section also included the gas, steam and hot water distribution activity in industries or residential buildings. This indicator, therefore, includes activities for electric and gas utilities, which generate, control and distribute electric power or gas. It also included steam and air-conditioning system supplies. We select the companies that are engaged in providing service:

1. Electric power generation, transmission, and distribution
2. Manufacture of gas and distribution of gaseous fuels
3. Steam and air conditioning supply

### 3.3.5 Water supply, sewerage, waste management and remediation activities

This indicator includes activities related to the management of various types of waste such as industrial, household and other waste. Water collection, treatment and supply activities, sewage treatment, waste collection, treatment and disposal activities; remediation activities and other waste management services have been included this section. We select the companies that are engaged in providing service:

1. Water collection, treatment, and supply
2. Sewerage
3. Waste collection, treatment and disposal activities
4. Remediation activities and other waste management services

### 3.3.6 Construction

This indicator includes general construction such as full housing, office buildings, stores, and other government and utility buildings, farm buildings, and specialized construction activities for buildings and civil engineering work like motorways, railways, airfields, roads, bridges, tunnels, ports and other water projects, sewer systems, irrigation systems, industrial facilities, pipelines, and sports facilities, electric lines, etc. These include new works, repairs, additions and alterations, the construction of pre-built buildings or structures on the site and the construction of a temporary nature. This work can be carried out on your own account or on a fee or contract basis. Parts of the work and sometimes even the whole practical work can be subcontracted. Building and engineering repairing works are also included. We select the companies that are engaged in providing service:

1. Construction of buildings
2. Civil engineering
3. Specialized construction activities

### 3.3.7 Wholesale and Retail Trade, Repair of Vehicles and Motorcycles

This indicator includes wholesale and retail sales of any type of products and the relevant rendering in the sale of these products. Wholesale is the resale of goods retailers, to industrial, institutional, professional users, commercial, to other wholesalers. Retail

sales are basically the sale of new and used products to the general public. The retail product has the sale in shops, department stores, stalls, private or family Mail-order homes, door-to-door sales to people, hawkers and paddlers, consumer cooperatives, auction house, etc.

We collect the wholesale and retail points which are engaged in providing the following service:

1. Wholesale trade
2. Retail trade
3. All kinds of points which engaged in wholesale or retail trade

The related organizations types has given in Appendix B.4.

### 3.3.8 Transportation and Storage

This indicator includes the economic units that engaged in passenger or freight transport, by rail, road, water, air and associated activities such as terminal and parking facilities, cargo handling, storage and renting the transport equipment with driver or operator. The postal and courier service are also included here.

We collect the economic unit's information which is engaged in providing the following service:

1. Land transport and transport via pipelines
2. Water transport
3. Air transport
4. Warehousing and support activities for transportation
5. Postal and courier activities

The organizations types has given in the appendix B.5.

### 3.3.9 Accommodation and Food Service

This indicator includes the economic units that engaged in short-stay accommodation for visitors and the other travelers, and food suitable for immediate use and includes the provision of drinks. This section includes short term accommodation activities, camping grounds, recreational vehicle parks, and trailer parks, other accommodation services,

food, and restaurants, beverage service activities and event catering, mobile food service activities, and other food service activities, food and beverage delivery activities and so on.This section excludes long-term accommodation and food-drinks that are sold as wholesale or retail trade. We collect the economic unit's information which is engaged in providing the following service:

1. Accommodation
2. Food and beverage service activities

### 3.3.10 Information Services and Communication

This indicator includes the economic units that engaged in the production and distribution of information and cultural products, data communication, information technology, data processing, and other information service related activities. The activities such as telecommunications activities, software publishing, radio, and TV broadcasting, motion picture, and sound recording activities, programming activities and information technology are included in this section. We collect the economic unit's information which is engaged in providing the services as follows.

1. Publishing activities
2. Motion picture, video and television program production, sound recording and music publishing activities
3. Programming and broadcasting activities
4. Telecommunications
5. Computer programming, consultancy, and related activities
6. Information service activities

The organizations types has given in the appendix B.6.

### 3.3.11 Financial and Insurance Services

This indicator covers financial services activities, including insurance, banking, reinsurance, pension funding and activities to support financial services. This indicator also covers financial services activities like central banking, monetary inter-mediation, other financial intermediaries, activities of funds, holding companies, trusts and similar financial entities, financial leasing, auxiliary financial service activities, credit granting,

financial market administration, fund management activities and activities of insurance agents and brokers. According to BBS, the contribution of this indicator to GDP was 3.93% in fiscal year 2017-18.

We collect the economic unit's information which is engaged in providing the following service:

1. Financial service activities
2. Insurance, reinsurance and pension funding activities
3. Auxiliary financial service activities

The organizations types are given in Appendix B.7.

### 3.3.12 Real-Estate Services

This indicator covers the economic units that include selling or buying real estate, renting real estate, providing real estate valuation or other real estate services such as acting as a real estate escrow agent. The activities of this section may be conducted on its own or lease property and may be performed on a fee or contract basis. These include structural buildings, such as those associated with owning or leasing structures.

We collect the economic unit's information which is engaged in providing the following service:

1. The real estate activities with own or leased property
2. The real estate activities on a fee or contract basis

### 3.3.13 Professional, Scientific and Technical Activities

This indicator covers the economic units that include specialized professionals, scientific and technical activities. These types of activities require a high degree of training, and make specialized knowledge and skills available to user. This indicator includes legal and accounting activities, book keeping and auditing activities; architectural and engineering activities, technical testing and analysis, tax consultancy activities, management consultancy activities, scientific research and development, research and experimental development on natural sciences and engineering, specialized design activities, advertising and market research, public opinion polling, veterinary activities, photographic activities and other professional, scientific and technical activities. We collect the information which is engaged in providing the following services.

1. Legal and accounting activities
2. Management consultancy activities
3. Architectural and engineering activities
4. Technical testing and analysis
5. Scientific research and development
6. Advertising and market research
7. Other professional, scientific and technical activities

The organizations types have been given in Appendix B.8.

### 3.3.14 Administrative and Support Services

This indicator covers the economic units that includes a variety of activities that support general business operations such as rental and leasing activities, leasing of intellectual property and similar products, travel agency, tour operator, reservation service and related activities, activities of employment placement agencies, activities of temporary employment agencies, other human resources provisions, maintenance service activities, services to buildings and landscape activities, general cleaning of buildings, combined facilities support activities, cleaning activities, office administrative, office support and other business support activities, security and investigation activities, security systems service activities, private security activities, investigation activities, other building and industrial cleaning activities, landscape care etc. We collect the economic units of information related to the following services:

1. Rental and leasing activities
2. Employment activities
3. Tour operator, travel agency, reservation service agencies and related service activities
4. Security and investigation activities
5. Services to buildings and landscape activities
6. Office administrative and office support activities

The organizations types has given in the appendix B.9.

### 3.3.15 Public Administration, Defence, Compulsory Social Security

This indicator covers the units that include the activities of a governmental nature, normally carried out by the public administration. This indicator includes administration of the state and the economic and social policy of the community, legislative activities, taxation, general public administration activities, defense activities, regulation of the activities of providing education, health care, cultural services and other social services, laws and judicial activities regulation of and contribution to more efficient operation of businesses, administrative programmes, provision of services to the community as a whole, foreign affairs, public order and safety activities, compulsory social security activities and compulsory social service activities etc. We collect information which is engaged in providing the following service:

1. The administration of the state and the economic and social policy of the community
2. General public administration activities
3. Regulation of the activities of providing health care, education, cultural services
4. Other social services
5. Regulation of more efficient operation of businesses
6. Foreign affairs
7. Defense activities
8. Public order and safety activities
9. Compulsory social security service activities

### 3.3.16 Education

This indicator covers the establishment that engaged in private or public education at any level or for any profession. The education may be oral or written as well as by radio and television or other any kind of communication. It includes the education by the different establishments in education system at its different levels as well as literacy programs, special education for physically or mentally handicapped pupils, adult education, religious education, computer education, technical education, training centres, professional coaching centres, military schools and academies, prison schools etc. We collect the establishment information which is engaged in providing the following education services.

1. Pre-primary and primary education
2. Secondary education
3. General secondary education
4. Technical and vocational secondary education
5. Higher education
6. Sports and recreation education
7. Cultural education
8. Educational support activities
9. Other education

### 3.3.17 Human Health and Social Work

This indicator covers the establishment that involves in health and social work activities. This includes hospital, clinic, or any kind of health care providing organization, and the organizations those have provided social service activities.

We collect the establishment information which is engaged in providing the following service:

1. Human health activities
2. Residential care activities
3. Social work activities without accommodation

### 3.3.18 Arts AND Entropy

The indicator covers the establishment that includes a wide range of activities to cater to the diverse cultural, recreational and recreational interests of the general public, including live performances, management of museum sites, activities of amusement parks and theme parks, sports and recreation activities, botanical and zoological parks and nature reserves activities.

We collect the establishment information which is engaged in providing the following service:

1. Creative, arts and entertainment activities
2. Libraries, archives, museums and other cultural activities
3. Sports activities and amusement and recreation activities

### 3.3.19 Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use

This indicator covers the activities of household such as house maids, waiters, cooks, valets, laundresses, butlers, chauffeurs, gatekeepers, stables-lads, caretakers, tutors, secretaries, governors, babysitters etc. It also includes the undifferentiated subsistence goods-producing and services producing activities of households. This section includes the activities as follows:

1. Household activities as employers of domestic workers
2. Different types of goods and service production activities of private families for their own use

### 3.3.20 Activities of extraterritorial organizations and bodies

This indicator includes activities of international organizations such as the United Nations and the specialized agencies of UN, regional bodies etc., the IMF, the World Bank, the World Customs Organization, the Organization for Economic Co-operation and Development, the OPEC, the European Communities, the European Free Trade Association and other international organizations. This indicator also includes the activities of diplomatic and consular missions.

We collect the establishment information which is engaged as the following organization:

1. Activities of extraterritorial organizations and bodies

### 3.3.21 Other service activities

This indicator includes the activities that were not covered elsewhere in the classification. It includes the activities of membership organizations, repairing of computers and personal and household goods, funeral and related activities, washing and cleaning of textile and fur products, hairdressing and other beauty treatment, and last of all, a variety of personal service activities (e.g., student, housewife) not covered elsewhere in the classification. We collect the information of the organization or persons which is involved with the following activities:

1. Activities of membership organizations
2. Repairing of computers and personal and household goods

3. Other personal service activities

The organizations types has given in the appendix B.10.

## 3.4 Occupation Finding

Occupation means an activity or work with which one occupies himself; especially productive activities, services, commerce, or craft for which one is regularly paid; Job. It is a person's regular work or profession. Occupation can be any work like engineering, government job, a family business, actors, business man, doctors, clerk, banker, teacher, singer, writer etc. Occupation provide us fulfill our basic needs-food, cloth and shelter. We select that user who shares his/her occupation in social media. Our goal is count the number of employees of each city. There are four indicators of Occupation Dimension as given in the table 3.2.

| ***Dimension*** | ***Indicators*** |
|---|---|
| **Occupation** | 1. Employment<br>2. Business<br>3. Non-Residents<br>4. Other Occupation |

Table 3.2: Defining indicators of *Occ.* dimension.

A sample scenario of city ranking at *occupation* dimension has given in the figure 3.5.

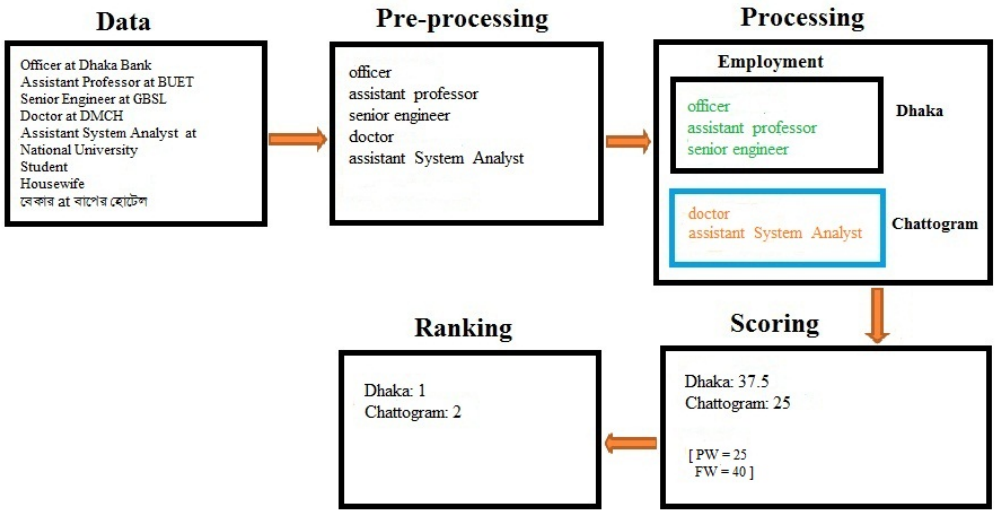



Figure 3.5: A sample scenario of geo-financial ranking at Occupation dimension.

### 3.4.1 Employment

Employment is a relationship between two parties namely, the employer and the employee. It is usually based on a contract where the work is paid for in the profitable or non profitable organization or in the company. Employees work for payments which can be in the form of hourly, weekly, monthly or an annual pay, depending on the kind of work an employee is doing or in which sector he is working. Employees in some fields or sectors may receive gratuity, bonus payments or stock options depending on the nature of employer. In some types of jobs, employees can receive payment as well as some benefits such as health insurance, housing, disability insurance or use of gym etc. Employment is usually governed by employment law, regulation or legal contract.

The five main employment types are:

1. Permanent or fixed-term employees
2. Casual employees
3. Apprentices or trainees – employees
4. Employment agencies staff – known as labor hire
5. Contractors and sub-contractors – known as hired staff

An employment is a job such as teacher, doctor, engineer, banker etc. This indicator includes the professional designation that an employee holds. In this research, we count the number of employees for each city through social media with the aim of detecting city's financial solvency.

### 3.4.2 Business

Business is an activity of producing, buying and selling products to customers. There are three major types of business:

1. **The Service Business:** This type of business provides intangible products such as professional skills, expertise, advice, and other similar products to their customers. The examples of service businesses: salons, repair shops, accounting firms and law firms, schools, banks etc.
2. *Merchandising Business:* In merchandising business, products are buys at wholesale price and make profit by selling those products at higher price than purchase price. So, this types of business is known as "buy and sell" business. In this business, the product is selling without any change of its actual form. Examples of these businesses are: grocery stores, distributors, convenience stores, and other re-sellers.

3. *Manufacturing Business:* In a manufacturing business, a new product is produce from raw materials with using manufacturing process. In this business, the product is selling with the change actual form. This type of business use raw materials to produced a new product. Last of all, the manufactured products is selling to customers.

{Hybrid Business: A Hybrid business is a business where the companies may be classified in more than one type of business. A restaurant, for example, combines ingredients in making a fine meal (manufacturing), sells cold water (merchandising), and fills customer orders (service).

**Business ownership:** There are seven basic forms of business ownership's:

1. **Sole proprietorship business:** A business owned by a single person.
2. **Partnership business:** In a partnership business, the business is owned by two or more people who share responsibilities, profits and loss of business.
3. **Limited Partnership business:** General partner(s) run the business, while limited partners invest.
4. **Corporation:** A fully- independent business with shareholder.
5. **Limited Liability Company (LLC):** A mixture of a partnership and a corporation business which is designed to make easier to start small business.
6. **Non Profit Organization:** A business that uses its profits for charitable purposes.
7. **Cooperative (Co-op) business:** A business owned and operated for the benefit of the members of the organization that use it services.

### 3.4.3 Non Residents

A Non-Resident Bangladeshi (NRB) is a Bangladeshi citizen who holds a Bangladeshi passport and has temporarily emigrated to another country for work, residence or any other purpose and living abroad six months or more. The term "NRB" covers both the permanent and temporary of Bangladeshi origin– who are working or living abroad for a long time (six month or more). The permanent migrants living in different countries are known as the "Bangladeshi diaspora".
More than 10 million Bangladeshis are living abroad of which about 2.4 million Bangladeshis are living abroad permanently and the rest of living temporarily. They lives in 162 countries like Saudi Arabia, United Arab Emirates, Qatar, Kuwait, Lebanon,

Oman, Bahrain, Japan, Australia, Yemen, UK, USA, Italy, Greece, South Africa, France, Netherlands, Belgium, Canada, Spain, Germany, Switzerland and other countries.
The NRB sends remittance on a regular basis. Remittance is second highest source of foreign currency earnings of Bangladesh after exports of ready-made garments (RMG). The annual remittance of the country has increased significantly as the migrant workers from Bangladesh gradually increase over the years. So, remittance has become a major contributor to the Bangladeshi economy. In the year 2018, the expatriates have sent 15.54 billion of foreign currency. The contribution of foreign remittance sent by the NRB makes up about 12 percent of Bangladesh's GDP.

### 3.4.4 Other Occupation

This section includes other categories that are not included in previous indicators. Example: student, housewife, house maid etc.

## 3.5 Daily Financial Activities (DFA)

How much a person costs or earned daily? An event related to financial that occurred in our daily life and we share it with our friends in social media. For example: buying a new flat, traveling in Nepal, having dinner in a five star hotel. we add this dimension to find the financial transaction status of cities. We think this is more dominant factor to solvency analysis. Our target is count the number of event related to financial as city wise. There are three indicators of DFA Dimension. The table 3.3 shows the indicators of DFA dimension.

| ***Dimension*** | ***Indicators*** |
|---|---|
| **Daily Financial Activities** | 1. Buying or Purchasing<br>2. Eating and Drinking<br>3. Entertainment |

Table 3.3: Defining indicators of *DFA* dimension.

A sample scenario of Ranking at *DFA* dimension has given in the figure 3.6.

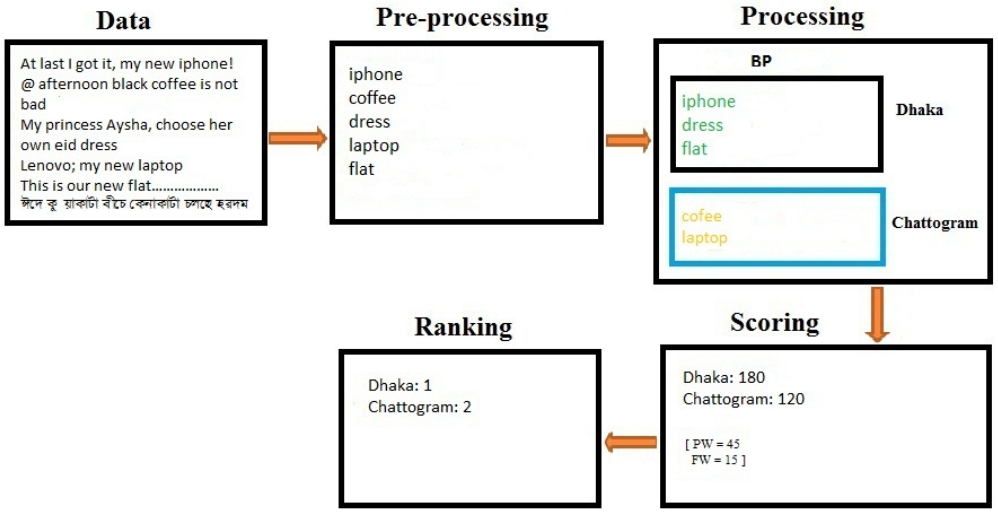



Figure 3.6: A sample scenario of geo-financial ranking at DFA dimension.

### 3.5.1 Buying or purchasing

Buying is most commonly used to refer to everyday goods and commodities, on the other hand, Purchasing is considered to be a more formal term than buy. Purchasing is often used to refer to contracts and large products, on the other hand, buying refers to smaller products.

For example, *'I purchased a new flat'. 'The government purchased a huge defense contract from the U.S. military'. However, 'I bought a new phone'.*
Still, it must be noted that in daily usage, the both terms (buy and purchase) are synonymous. For example, *I bought a new book'. I purchased a new book'.*

### 3.5.2 Eating and Drinking

This indicator includes all kinds of food and drinks that we eat daily or occasionally and share them with our friends on social media (e.g., Facebook). For example,*'With my best friend, Sajib, at Modhumitta Coffee shop'. 'Haji Biriani is so tasty'.*

### 3.5.3 Entertainment

The arts, entertainment, and recreation are comprised of a variety of organizations that provide benefits or services to their patrons to meet different cultural, entertainment and recreational interests. This index includes companies that are engaged in the production, promotion or participation in live performances, events or exhibitions intended for public viewing; the establishments that preserve and display historical, cultural or educational

objects and sites; and establishments that provide services or operate facilities that enable peoples to engage in recreational activities or pursue amusement, hobbies and leisure interests.

## 3.6 Data Collection

There are two types of data. (1) Economic Establishment data which has been collected from web media (Bangladesh Yellow Pages), and (2) Social media data which has been collected from Facebook.

### 3.6.1 Economic Establishment Data

Economic Establishment data has been collected from many web sites which collects company information of Bangladesh. We collect data through web scraping. We take bangladeshyellopages.com data. We use *R* scripts use for downloading data. The web page contains many kinds information, all are not required for us. We take those data which is needed for our research. Therefore, to select those type of data we installed the "Selecter Gadget" to Browser to select relevant data table. The data collection procedure is given as below:

1. Install Selector Gadget on google browser from https://selectorgadget.com/
2. Select the relevant data table of bangladeshyellopages.com using Selector Gadget
3. Using *R* script for downloading selected data
4. Run the R script on R studio
5. Then the unstructured data downloaded as text format and we saved it in a csv file

The figure 3.7 shows the data collection process using web media.

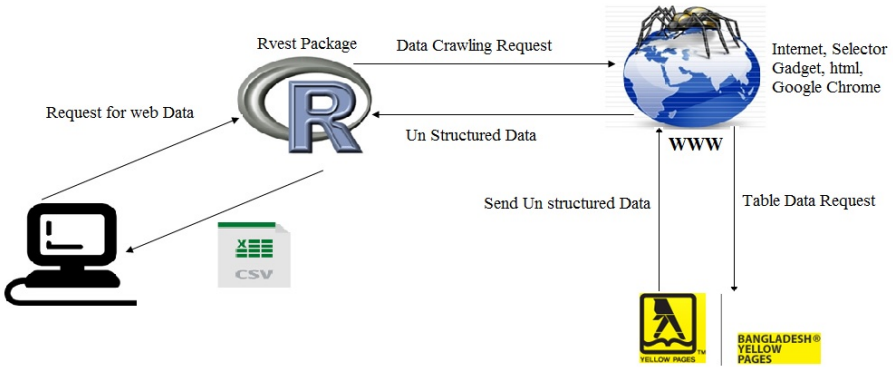



Figure 3.7: Economic establishment data collection.

### 3.6.2 Social media data collection

For Occupation and DFA data collection, we analyze social media posts. Firstly, we considered twitter, facebook, yahoo etc. as social media. However, mostly Bangladeshi people use Facebook. Therefore, we target facebook users for data collection. We collected data from user who permitted us to access their personal data. The figure 3.8 shows the data collection process using social media.

#### Personal Information

Users provide a wealth of personal information in his/her Facebook profile. We collected the raw data included features like the user's name, hometown, current city, profession, and workplace. From current city we detect them as the residents or non-residents. This include values of 0 for the users who did not supply any information.

#### Activities and Preferences

Providing lists of personal activities or favorite things have been always been a part of Facebook. We collected the user's posts which are related to financial activities like travel, buying something, about food, about entertainment. We counted the number of activities as city wise.
*R* scripts have been used for processing the posts or data.The following steps are taken for social media data collection:

1. To create an apps through https://developers.facebook.com/tools
2. To invite the facebook users to give permission to use on their data by registration on that app or by using graph API explorer.
3. To get the user access token
4. We write a *R* script for download the user data
5. Run the R script on R studio
6. Then the unstructured data downloaded as csv format

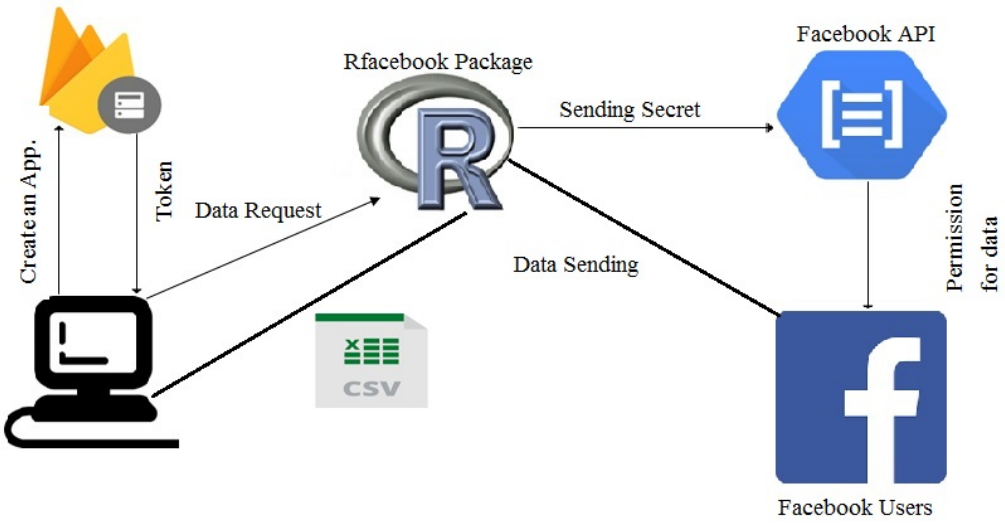



Figure 3.8: Social media data collection process

## 3.7 Data Pre-processing

*Data pre-processing* technique involves the conversion of raw data into an understandable format. Real-world data is usually incomplete, inconsistent, and/or is a lack of specific behaviors or tendencies and can have many flaws. The *Data pre-processing* technique is a proven method to resolve such flaws. There are many techniques [17–19] which have been studied for pre-processing the data. Data pre-processing steps involves: (1) citizen selection in social media, (2) Data cleaning.

### 3.7.1 Citizen selection from social media

To find the citizen of our targeted areas we performed the following steps:

1. Check the user home town with area dictionary.
2. Remove the user who is not the citizen of our targeted areas.

The figure 3.9 shows a scenario of citizen selection procedure in social media.

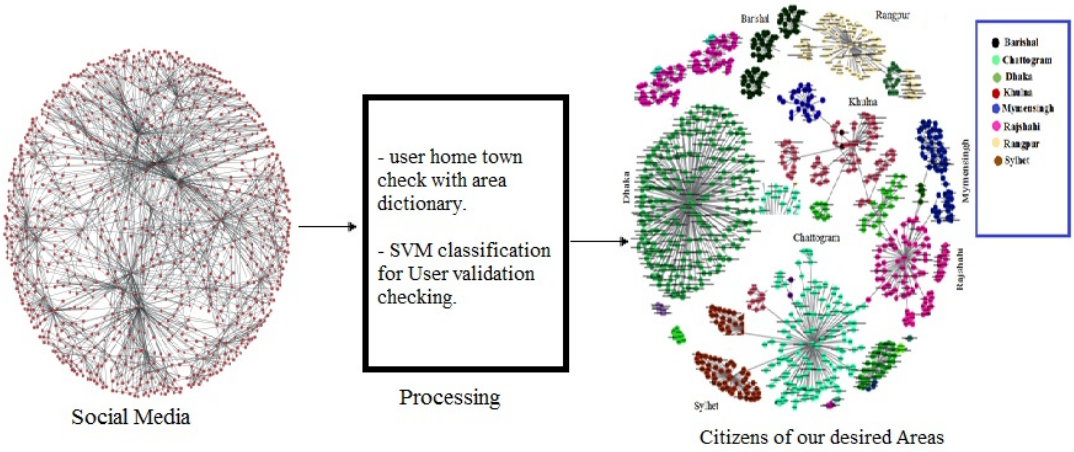



Figure 3.9: A scenario of citizen selection in social media.

### 3.7.2 Pre-processing of Economic Establishment Data

Pre-processing of establishment data includes the following tasks:

1. To make all text into lowercase using tolower() function.
2. To remove all blank lines using regular expression.
3. To remove all URLs, hash tags, targets using the function removePunctuation().
4. To removes tabs and extra spaces using stripWhiteSpace()
5. To remove unnecessary words using removeWords()
6. To remove Non-English lines

The figure 3.10 shows the steps of data pre-processing to get structured data.

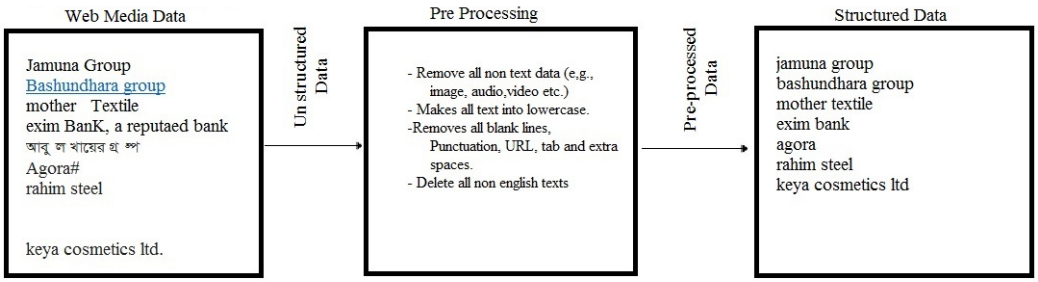



Figure 3.10: The pre processing steps of web media data at *EE* dimension

### 3.7.3 Pre-processing of Social Media Data

Pre-processing of social media data includes the following tasks:

1. To make all text into lowercase using tolower() function.
2. To remove all blank lines using regular expression.
3. To remove all URLs, hash tags, targets using the function removePunctuation().
4. To removes tabs and extra spaces using stripWhiteSpace()
5. To remove unnecessary words using removeWords()
6. To remove Non-English lines
7. to correct the spellings; sequence of repeated characters is to be handled.
8. to remove all punctuations, symbols, numbers
9. to remove Stop Words: There are some words that are frequently used in English sentences but provide little information. These are known as stop words, and we may want to remove them from our analysis. For example, "I", "she'll", "the", etc. There are 175 common English stop words are found in tm package which is given as below.

**STOP WORDS:**

you, your, yours, yourself, yourselves, we, our, ours, ourselves, i, i'm, me, my, myself, he, him, too, very, will, his, himself, she, her, hers, herself, it, its, itself, they, them, their, theirs, themselves, have, has, had, having, do, does, did, doing, what, which, who, whom, this, that, these, those, am, is, are, was, were, be, been, being, i've, you've, we've, they've, i'd, would, should, could, ought, you're, he's, she's, it's, we're, they're, you'd, he'd, she'd, we'd, they'd, doesn't, don't, didn't, i'll, you'll, he'll, she'll, we'll, they'll, isn't, aren't, wasn't, weren't, hasn't, haven't, hadn't, won't, that's, who's, what's, wouldn't, shan't, shouldn't, can't, cannot, couldn't, mustn't, let's, here's, there's, when's, where's, why's, of, at, by, for, with, how's, a, an, the, and, but, if, or, because, as, until, while, about, to, from, up, down, in, against, before, after, above, below, when, where, why, how, no, nor, not, out, on, off, over, under, again, all, between, into, through, during, any, both, each, few, more, most, other, than, further, then, once, here, there, some, such, only, own, same, so.

## 3.8 Data Processing

Data processing has been done for manipulating the data indicator wise and to make data fit for score calculation. There are many techniques [20–23], have been studied for data processing. In this research the data processing is done with the following tasks:

1. User validation checking
2. Dictionary construction
3. Financial word counting from user posts
4. Number of non-residents counting
5. Number of employee counting

### 3.8.1 User validation checking

To detect valid user of social media we perform the following steps:

1. We use SVM classifier for User validation checking
2. We take those users who provide at least three types of information in his/her facebook profile : home town, location/Current City and work/profession.
3. We remove those users who post under twenty [24], posts to discourage insufficient information for profile computing.
4. SVM provides the best classification accuracy and therefore, we take SVM for social media user validation checking. (Classification is done with WEKA software.)

The figure 3.11 shows the flow Chart for valid user checking in social media.

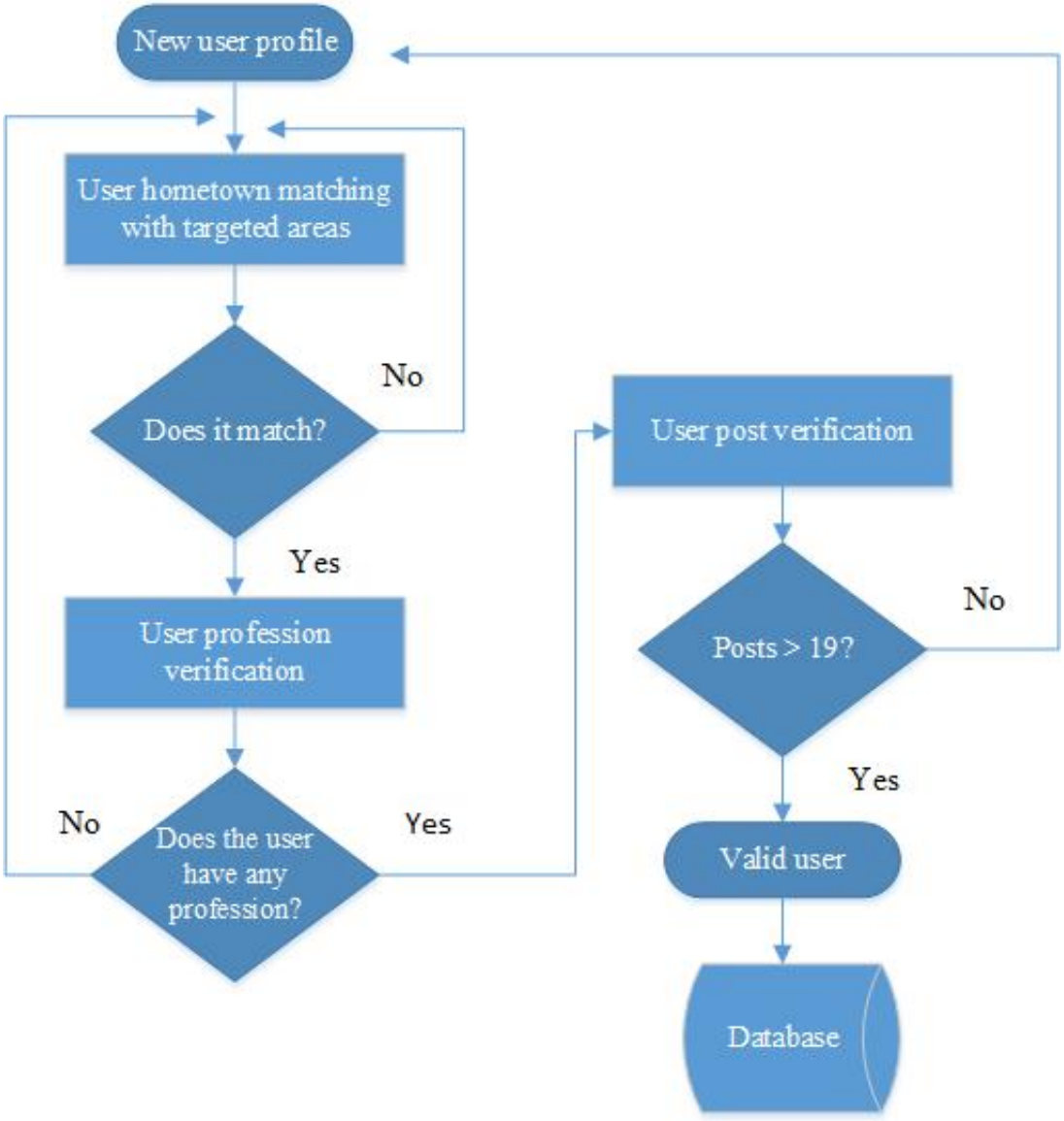



Figure 3.11: The flow chart of citizen selection procedure in social media.

The table 3.4 shows SVM classification to select valid social media user.

| Id | Hometown | Location | Work | Posts | Valid |
|---|---|---|---|---|---|
| 10214826143312460 | 1 | 1 | 1 | 20 | Yes |
| 558523984664452 | 0 | 1 | 1 | 31 | Yes |
| 104725000622012 | 0 | 0 | 0 | 21 | No |
| 664032562182458 | 1 | 1 | 1 | 8 | No |

Table 3.4: A sample of data classification procedure to detect valid citizen in social media

The user provides information=1 and the user provides no information= 0

#### Using different Classifiers

We use different classifier like linear regression, voted perception and SMO to check user validity. Among three the SVM performs better, so, we use it for social media user selection.

### 3.8.2 Financial word counting from user's posts

We constructs four dictionaries as follows:

1. **Area dictionary:** It contains the name of our 8 divisional cities for citizen selection.
2. **Word dictionary:** It contains financial related words for counting financial word from user posts.
3. **Country dictionary:** It contains all countries and famous cities name in the world for finding non-residents. But it excludes the target areas name (eg., in this research, Bangladesh and the eight divisional cities name).
4. **Occupation dictionary:** It contains the designation name of different profession.

The figure 3.12 describes the steps of financial word counting from user's posts.

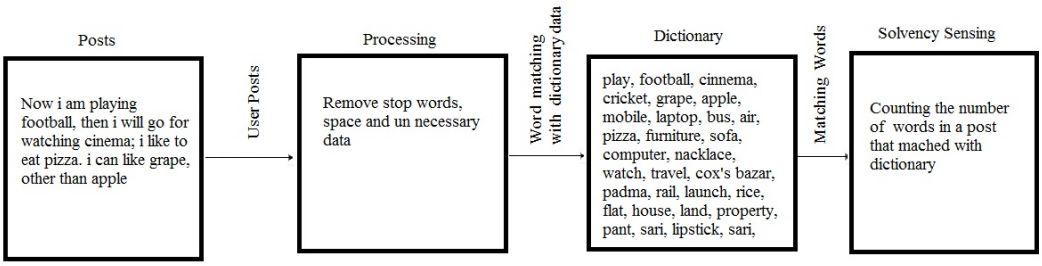



Figure 3.12: The steps of finding financial related word from user's post in social media.

### 3.8.3 Non-residents computation

The objective of this section is counting the number of non-residents by matching the user location with country dictionary. The figure 3.13 depicts the steps of non-residents detection from user's profile.

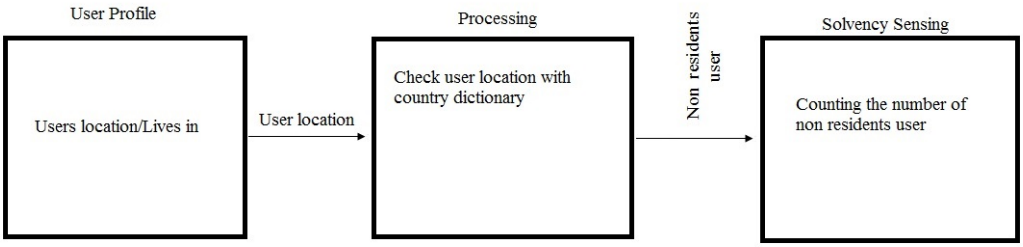



Figure 3.13: The steps of finding Non residents from social media users.

### 3.8.4 Employee finding

The objective of this section is counting the number of employees in an area by ignoring non-employed users. When we get a user's profession we match it to occupation dictionary. If it matching we just count the number of employees, businessman as city wise. In the case of user who does not provides the information of designation, we just count there the number of organization name. The figure 3.14 express the steps of employee finding.

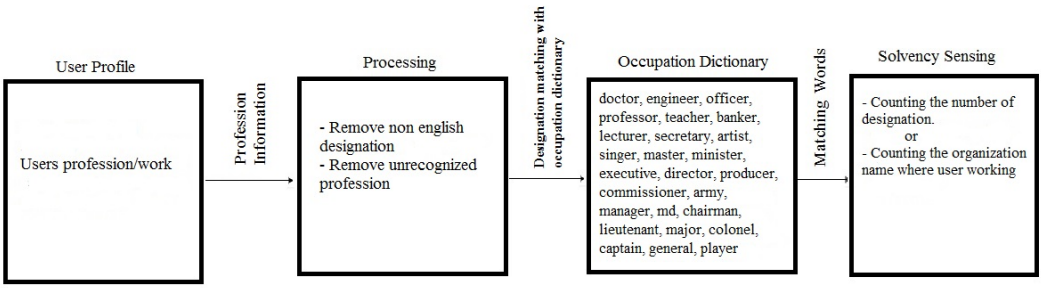



Figure 3.14: Finding employees from user profession

## 3.9 Score Calculation

The purpose of this section is to calculate dimensions and indicators scores for each city. There are two steps for calculating score which are noted as follows:

1. **Weight Distribution:** which describes the distribution of primary and final weight among dimensions and indicators.

2. **Scoring on *EE*, *Occ.* and *DFA* dimension:** which describes the score calculation of the areas based of proposed method.

### 3.9.1 Weight Distribution

The objective of this section is to assign weight for each dimension and indicators. To achieve this goal, we assign two types of weight (1) Primary Weight and (2) Final Weight.

- The Primary Weight, PW, is assigned to 28 indicators to measure the indicator level score.
- The Final Weight, FW, is assigned to 3 dimensions to measure the dimension level score.

#### Weight assigning at Dimension level

To assign weight among dimensions, we have studied many national and international standard methods [1,8,16]. We called the dimension level weight is the Final Weight (FW). The final weight of 3 dimensions ( *EE, Occ., DFA*) has distributed on a scale of 1-100. The table 3.5 describes the weight distribution statistics among three dimensions.

| Dimension | FW |
|---|---|
| **Economic Establishment** | 45 |
| **Occupation** | 40 |
| **Daily Financial Activities** | 15 |
| **Total** | 100 |

Table 3.5: Distribution of weight (FW) at dimension level

#### Weight distribution to the indicators of EE dimension

The national contribution of 21 indicators in Bangladeshi economy finding by economic census-2013 [1], is considered as primary weight in economic establishment indicators. The table 3.6 describes the primary weight distribution statistics among 21 indicators of EE dimension.

| S/N | Indicator Name | PW |
|---|---|---|
| 1. | AFF | 2.9 |
| 2. | MAQ | 0.22 |
| 3. | MFG | 11.76 |
| 4. | EGSA | 0.15 |
| 5. | WSWRA | 0.75 |
| 6. | Const. | 2.39 |
| 7. | WRT | 45.91 |
| 8. | TAS | 13.65 |
| 9. | AFSA | 4.77 |
| 10. | IC | 0.55 |
| 11. | FIA | 0.47 |
| 12. | REA | 0.15 |
| 13. | PSTA | 0.65 |
| 14. | ASSA | 0.39 |
| 15. | PAD | 0.10 |
| 16. | Edu. | 3.93 |
| 17. | HHSWA | 2.03 |
| 18. | AER | 0.39 |
| 19. | AHE | 0.34 |
| 20. | AEOB | 0.03 |
| 21. | OSA | 8.48 |
| | TOTAL | 100 |

Table 3.6: Primary weight distribution on indicators of *EE* Dimension

### Weight distribution to the indicators of *Occ.* dimension

The table 3.7 describes the primary weight distribution statistics among 4 indicators of Occ. dimension. The primary weight of the indicators of *Occ.* dimension has calculated on a scale of 1-100.

| S/N | Indicator Name | PW |
|---|---|---|
| 1. | Emp. | 25 |
| 2. | Bus. | 30 |
| 3. | NR | 35 |
| 4. | OO | 10 |
| | TOTAL | 100 |

Table 3.7: Primary weight distribution on indicators of *Occ.* Dimension

### Weight distribution to the indicators of DFA dimension

The table 3.8 describes the primary weight distribution statistics among 3 indicators of DFA dimension. The primary weight of the indicators of *DFA* dimension has calculated on a scale of 1-100.

| S/N | Indicator Name | PW |
|---|---|---|
| 1. | BP | 45 |
| 2. | ED | 30 |
| 3. | Ent. | 25 |
| | TOTAL | 100 |

Table 3.8: Primary weight distribution on indicators of *DFA* Dimension

### 3.9.2 Score Computation for EE, Occ. and DFA

Let, $X = x_1, x_2, x_3, ....., x_n$ is the name of areas. In this research we have eight areas, $X = Barishal, Chattogram, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, Sylhet$. Let, $D = d_1, d_2, d_3, ....., d_n$ is the number of dimensions. In this work we have three dimesions, $D = EE, Occ., DFA$.
Let, $I = i_1, i_2, i_3, ....., i_n$ is the number of indicators. In this work we have total twenty eight indicators, $I = AFF, MAQ, Emp., ....., Ent.$.

We have six steps for score calculation at dimension and indicators levels. We repeat these steps for each city at each indicator and dimension level. At first, we count the numbers of Economic Establishments, Occupation and DFA data for each city using Equation 3.1.

$$S_i = \sum_{i=1}^{n} L_i \tag{3.1}$$

Where $S$= total number of counted data at each indicator level, $n$ = total number of indicators in a dimension (For EE dimension, n=21; for Occupation n=4, and for DFA, n=3), and $L$=number of establishments, posts, employees for each indicator of a city.
For example, in this research, we counted 120 manufacturing companies in Dhaka city at *EE* dimension. We get 6 employees in social media in *Occ.* dimension and we counted 192 buying and purchasing activities at *DFA* dimension.

Let, the counted number of EE, Occ. and DFA data are stored in the data table A. The A has M rows and N. where M is the row of the table for EE, M=21, for Occ., M=4 and for DFA, M=3 where N= the number areas, here N is constant and its value is 8 as we have 8 cities. Now, we will make average the data as row wise. Point Calculation (row level averaging) is as follows:

$$Avg_i = (\frac{S_i}{T_i}) * 100 \tag{3.2}$$

Where T= Total counted number of data as row wise. $T_i = \sum_{i=1}^{n} S_i$
The equation 3.2 shows the row level averaging of counted data.

Now, we will multiply the row level average value with our distributed weight. Firstly, multiply Avg value with primary weight at indicator level and secondly, multiply Avg

value with the final weight at dimension level.

$$AW_i = Avg_i * weight \tag{3.3}$$

The equation 3.3 shows the multiplication of row level average value with distributed weight. Now, we will find the Total point value by performing column wise summation of data table A.The equation 3.4 shows the column wise summation at indicator and dimension level.

$$Point_x = \sum_{i=1}^{l} AW_l \tag{3.4}$$

Where $l$ = the number of indicators mentioned in [3, 4, and 21] for three dimension levels.

To compute the score, we divide the total point value by total weight (100) at indicator and dimension wise. The equation 3.5 shows the score calculation.

$$Score_x = \frac{Point_x}{T_w} \tag{3.5}$$

Where $T_w$ is total weight per dimension level

Now, we will compute the Final score for each city by adding the score of three dimensions.

$$FScore_{x1} = (Scored1_{x1} + Scored2_{x1} + Scored3_{x1}) \tag{3.6}$$

The equation 3.6 shows the Final score calculation.

## 3.10 Computation of Rank

*Rank* is calculated for each dimension level based on weighted score value. We use R built-in-function to compute the rank at each dimension level. The equation 3.7 shows the rank computation at each dimension level.

$$Rank_X = (Score_{x1}, Score_x2, Score_{x3}.....Score_{xn}) \tag{3.7}$$

Final rank is computed over the final score of each city's in three dimensions occupation, EE, DFA. We uses R built-in-function to compute the final rank of a city. Equation 3.8 computes the final ranking of divisional cities. .

$$FRank = rank(FScore_{x1}, FScore_{x2}, FScore_{x3}, ....., FScore_{xn}) \tag{3.8}$$

# Chapter 4

# Results and Analysis

In this chapter, the experimental results and analysis are reported. At first, we provide the description of testing environment.

## 4.1 Tools and Techniques

We use following tools for processing our data.

1. R software: RStudio-1.1.383.
2. Weka 3.8
3. Classification method: SVM

The computation has been processed in a machine with Intel i3-6100 CPU, 3.70 GHz, RAM: 4.00 GB.

## 4.2 Results

In this section, we report our findings. We report the weighted score found for Occupation, Economic Establishment and Daily Financial Activities.

### 4.2.1 Economic Establishment

Here are the indicators of Economic Establishment of eight cities reported in Table 4.1.

| Indicator | Barishal | Chattogram | Dhaka | Khulna | Mymensingh | Rajshahi | Rangpur | Sylhet | Total |
|---|---|---|---|---|---|---|---|---|---|
| **AFF** | 15 | 64 | 90 | 31 | 13 | 12 | 18 | 21 | 264 |
| **MAQ** | 3 | 9 | 13 | 4 | 2 | 2 | 1 | 7 | 41 |
| **MFG** | 11 | 65 | 120 | 18 | 12 | 15 | 15 | 18 | 274 |
| **EGSA** | 9 | 63 | 180 | 18 | 7 | 19 | 12 | 37 | 345 |
| **WSWRA** | 5 | 32 | 131 | 18 | 17 | 18 | 12 | 14 | 247 |
| **Const.** | 15 | 67 | 141 | 39 | 14 | 18 | 16 | 24 | 334 |
| **WRT** | 46 | 152 | 281 | 38 | 36 | 39 | 52 | 32 | 676 |
| **TAS** | 95 | 160 | 300 | 91 | 90 | 75 | 60 | 85 | 956 |
| **AFSA** | 40 | 95 | 160 | 70 | 60 | 60 | 55 | 98 | 638 |
| **IC** | 12 | 45 | 81 | 48 | 10 | 14 | 11 | 20 | 241 |
| **FIA** | 30 | 48 | 85 | 36 | 27 | 31 | 26 | 51 | 334 |
| **REA** | 16 | 50 | 110 | 19 | 12 | 14 | 9 | 47 | 277 |
| **PSTA** | 8 | 31 | 67 | 21 | 8 | 8 | 13 | 12 | 168 |
| **ASSA** | 7 | 18 | 28 | 9 | 6 | 8 | 6 | 11 | 93 |
| **PAD** | 12 | 22 | 57 | 19 | 13 | 14 | 13 | 15 | 165 |
| **Edu.** | 14 | 19 | 55 | 20 | 12 | 18 | 9 | 11 | 158 |
| **HHSWA** | 20 | 53 | 110 | 24 | 19 | 21 | 14 | 18 | 279 |
| **AER** | 9 | 14 | 47 | 8 | 7 | 10 | 11 | 16 | 122 |
| **AHE** | 20 | 24 | 78 | 27 | 16 | 18 | 10 | 21 | 214 |
| **AEOB** | 4 | 12 | 32 | 6 | 5 | 5 | 3 | 6 | 73 |
| **OSA** | 25 | 37 | 119 | 28 | 26 | 24 | 16 | 19 | 294 |
| **TOTAL** | **416** | **1080** | **2285** | **592** | **412** | **443** | **382** | **583** | **6193** |

Table 4.1: Statistics on the data obtained in the indicators of *EE* Dimension

### 4.2.2 Demographic Data from Social Media

We collect data of ninety users of Facebook, from them we extract full data of 41 users and partial data of 49 users. From user profile we extract name, id, profession, hometown, current city and user posts.
*Full Dataset:* We call it *Full Dataset* when we extract name, id, profession, hometown, current city and posts.
*Partial Dataset:* We collect 49 user data from non-resident Bangladeshis. In this case, we collect name, id and current city. We call it *Partial Dataset*.

| City | TNU | TNP | TNNR |
|---|---|---|---|
| **Barishal** | 7 | 240 | 1 |
| **Chattogram** | 14 | 360 | 5 |
| **Dhaka** | 13 | 420 | 2 |
| **Khulna** | 9 | 251 | 1 |
| **Mymensingh** | 11 | 191 | 1 |
| **Rajshahi** | 8 | 260 | 2 |
| **Rangpur** | 6 | 185 | 2 |
| **Sylhet** | 22 | 521 | 13 |
| **Total** | **90** | **2428** | **27** |

Table 4.2: Statistics of demographic data of social media users.

Figure 4.2 depicts the social media data which we are collect from *Facebook*. Here, TNU denotes Total Number of Users, TNP denotes Total Number of Posts, and TNNR denoptes Total Number of Non-residents.

#### Occupation

Table 4.3 shows the statistics for different cities for *Occupation* dimension.

| City | Emp. | Bus. | NR | OO | Total |
|---|---|---|---|---|---|
| **Barishal** | 3 | 1 | 1 | 2 | 7 |
| **Chittagong** | 2 | 4 | 5 | 3 | 14 |
| **Dhaka** | 6 | 4 | 2 | 1 | 13 |
| **Khulna** | 4 | 2 | 1 | 2 | 9 |
| **Mymensingh** | 3 | 3 | 1 | 4 | 11 |
| **Rajshahi** | 2 | 3 | 2 | 1 | 8 |
| **Rangpur** | 1 | 1 | 2 | 2 | 6 |
| **Sylhet** | 2 | 6 | 13 | 1 | 22 |
| **Total** | **23** | **24** | **27** | **16** | **90** |

Table 4.3: Data obtained for the indicators of Occupation.

**Daily Financial Activities**

Table 4.4 shows the statistics of cities for DFA.

| City | BP | ED | Ent. | Total |
|---|---|---|---|---|
| **Barishal** | 44 | 111 | 85 | 240 |
| **Chattogram** | 109 | 98 | 153 | 360 |
| **Dhaka** | 192 | 153 | 75 | 420 |
| **Khulna** | 75 | 87 | 89 | 251 |
| **Mymensingh** | 67 | 69 | 55 | 191 |
| **Rajshahi** | 58 | 89 | 113 | 260 |
| **Rangpur** | 53 | 47 | 85 | 185 |
| **Sylhet** | 120 | 231 | 170 | 521 |
| **Total** | **718** | **885** | **825** | **2428** |

Table 4.4: Statistics on the data obtained in the indicators of *DFA* Dimension

## 4.3 Classifier Performance

We tested social media user data on three classifiers (1). Linear Regression (LR) (2). Voted Perceptron (VP) and (3). SVM The classification results have been reported as follows.
*Evaluation Metrics:*
We utilize four metrics for evaluating classifier performance to find the validity social media users in our research.

1. *Accuracy:* The accuracy of classification is defined as the number of correctly classified patterns to the total number of patterns. It can also be expressed as the ratio of the sum of true positives and true negatives to the total number of trials.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} = \frac{(Citizens\ Correctly\ Identified+Non\ Citizens\ are\ Correctly\ Identified)}{Total\ Number\ of\ Trials}$$

2. *Precision:* Precision tells us about the success probability of making correct positive class classification. It is computed as the number of True Positives divided by the total number of positive calls.
$$Precision = \frac{TP}{(TP+FP)} = \frac{Citizens\ Correctly\ Identified}{(Citizens\ Correctly\ Identified+Individuals\ Incorrectly\ Labeled\ as\ Citizens)}$$

3. textitRecall: Recall explains how sensitive the model is towards identifying the positive class. This is computed as the total number of True Positives divided by the total sum of True Positives and False Negatives.
$$Recall = \frac{TP}{(TP+FN)} = \frac{Citizens\ Correctly\ Identified}{(Citizens\ Correctly\ Identified+Citizens\ Incorrectly\ Labeled\ as\ Non\ Citizens)}$$

4. *F1 Score:* F1 score compute the score by considering both precision and recall of the test and It is the harmonic mean of the precision and recall. It reaches its best value

1 (perfect precision and recall) and worst at 0. The F1 score cannot be greater than accuracy.

$$F1\ Score = 2 * (\frac{(Recall * Precision)}{(Recall + Precishion)})$$

| | Metrics | | | |
|---|---|---|---|---|
| **Classifier** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| **LR** | 0.5 | 0.7 | 0.47 | 0.56 |
| **VP** | 0.54 | 0.01 | 1 | 0.02 |
| **SMO** | 0.58 | 0.78 | 0.53 | 0.63 |

Table 4.5: Classifiers to detect citizens from social media.

From Table 4.5, we have found that SMO performs better than other three classifiers. Therefore, we have selected SMO for data classification. The corresponding performance chart has given in the figure 4.1.

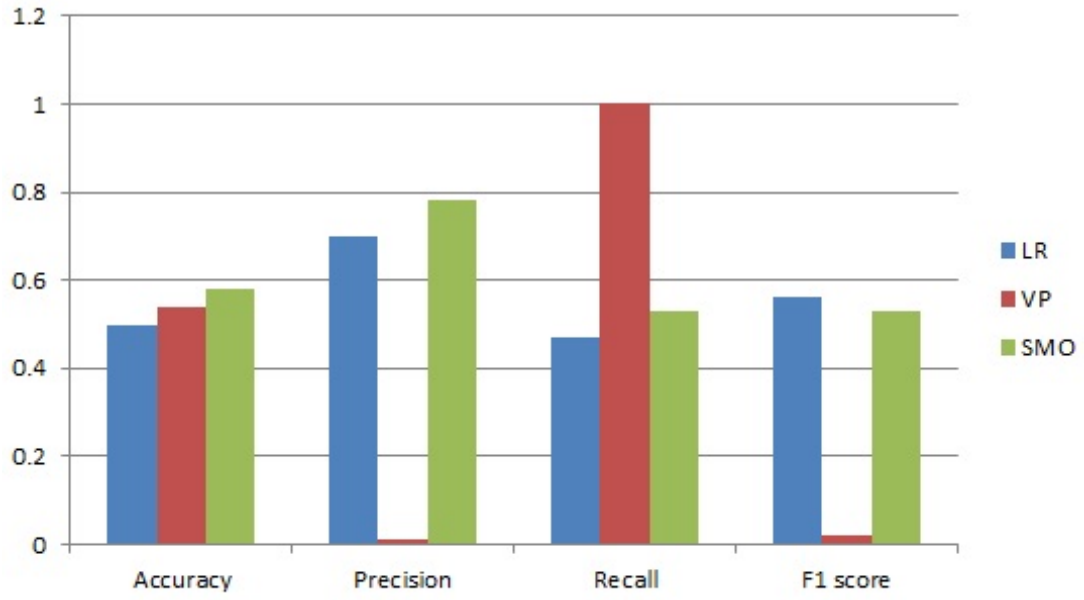



Figure 4.1: Classifier performance chart

## 4.4 More Results

From the analysis we have found Barishal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet score 6.98, 19.88, 38.92, 7.73, 6.37, 6.66, 6.79 and 6.67 respectively. Here, Dhaka have got the highest score and Mymensingh scored lowest. The table 4.6 describes the score of cities at *EE* dimension.

| City | Score |
|---|---|
| **Barishal** | 6.98 |
| **Chattogram** | 19.88 |
| **Dhaka** | 38.92 |
| **Khulna** | 7.73 |
| **Mymensingh** | 6.37 |
| **Rajshahi** | 6.66 |
| **Rangpur** | 6.79 |
| **Sylhet** | 6.67 |

Table 4.6: Score for cities in *EE* dimension.

### 4.4.1 Occupation Dimension

From the analysis we have found Barishal, Chattogram, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhetś scores are 7.06, 15.53, 14.74, 9.39, 10.81, 9.14, 6.18 and 27.15 respectively. Here, Sylhet has got the highest score, and Rangpur has scored the lowest. The table 4.7 expresses the score of cities at *Occ.* dimension.

| City | Score |
|---|---|
| **Barishal** | 7.06 |
| **Chattogram** | 15.53 |
| **Dhaka** | 14.74 |
| **Khulna** | 9.39 |
| **Mymensingh** | 10.81 |
| **Rajshahi** | 9.14 |
| **Rangpur** | 6.18 |
| **Sylhet** | 27.15 |

Table 4.7: The achieved score value of cities at *Occ.* dimension.

### 4.4.2 Scoring in *DFA* dimension

In DFA dimension, we have found Barishal, Chattogram, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhetś scores 9.1, 14.79, 19.49, 10.35, 8.2, 10.08, 7.49 and 20.5 respectively. Here, Sylhet has got the highest score, and Rangpur has scored the lowest. The table 4.8 shows the score of cities at *DFA* dimension.

| City | Score |
|---|---|
| **Barishal** | 9.1 |
| **Chattogram** | 14.79 |
| **Dhaka** | 19.49 |
| **Khulna** | 10.35 |
| **Mymensingh** | 8.2 |
| **Rajshahi** | 10.08 |
| **Rangpur** | 7.49 |
| **Sylhet** | 20.5 |

Table 4.8: The achieved score value of cities at *DFA* dimension

### 4.4.3 Overall Score Computation

The overall score is the summation of three dimensionsś scores. In this research We found Barishal, Chattogram, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet scores 7.33, 17.38, 26.33, 8.79, 8.42, 8.17, 6.65 and 16.94 respectively. Here, Dhaka has got the highest score and Rangpur has scored the lowest. The table 4.9 express the overall score of all cities in three dimension.

| City | Score |
|---|---|
| **Barishal** | 7.33 |
| **Chattogram** | 17.38 |
| **Dhaka** | 26.33 |
| **Khulna** | 8.79 |
| **Mymensingh** | 8.42 |
| **Rajshahi** | 8.17 |
| **Rangpur** | 6.65 |
| **Sylhet** | 16.94 |

Table 4.9: Final scores for all cities as measured by *SSM*.

## 4.5 Rank Computation

This section shows the ranks of the cites based on three dimensions, and the last of all we report the overall ranking.

### 4.5.1 Rank in *EE*

In EE dimension, we have found Barishal, Chittagang, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet ranks 4, 2, 1, 3, 8, 7, 5 and 6 respectively. Here, Dhaka got

the 1st rank and Barisal has got the 8th position. The graph 4.2 expresses the rank of the cities at *EE* dimension.

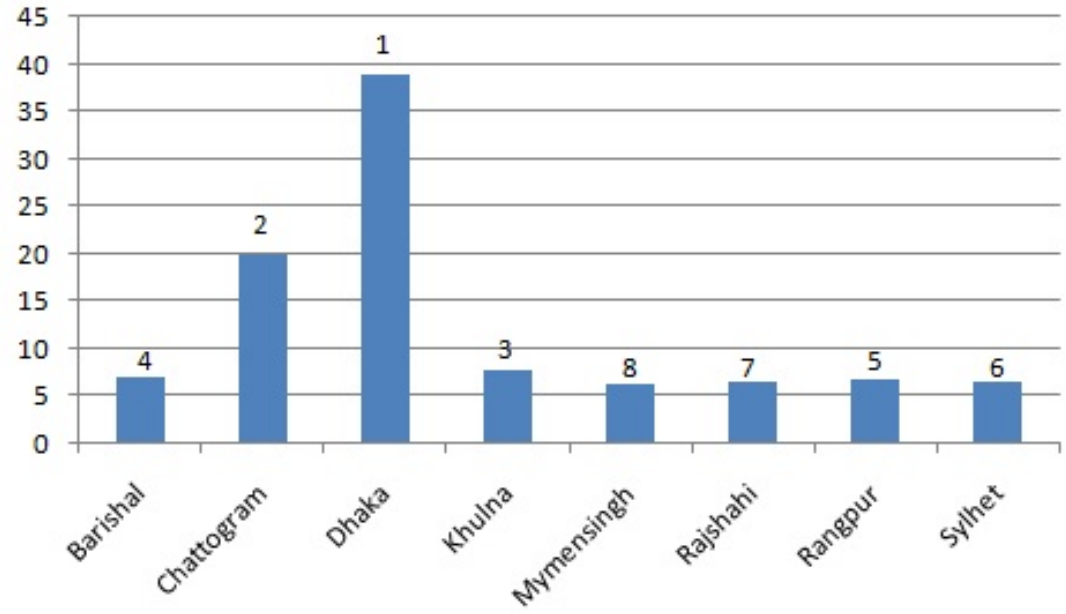



Figure 4.2: City rank in *EE* dimension

### 4.5.2 Rank in *Occupation*

In *Occupation*, we have found Barishal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet ranks 7, 2, 3, 5, 4, 6, 8 and 1 respectively. Here, Sylhet has got the 1st rank and Rangpur has got the 8th rank. The graph 4.3 describes the rank of cities at *Occupation*.

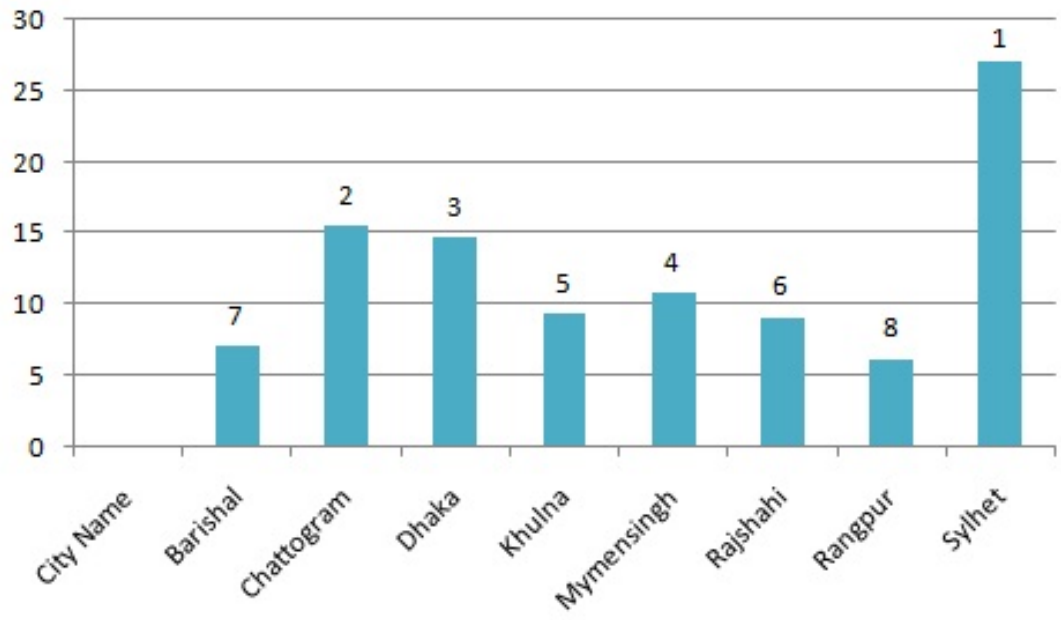



Figure 4.3: City rank in *Occupation*.

### 4.5.3 Ranking in *DFA*

In DFA, we have found Barishal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet's ranks as 6, 3, 2, 4, 7, 5, 8 and 1 respectively. Here, Sylhet got the 1st rank and Rangpur got the 8th. The graph 4.4 shows the rank of cities for *DFA*.

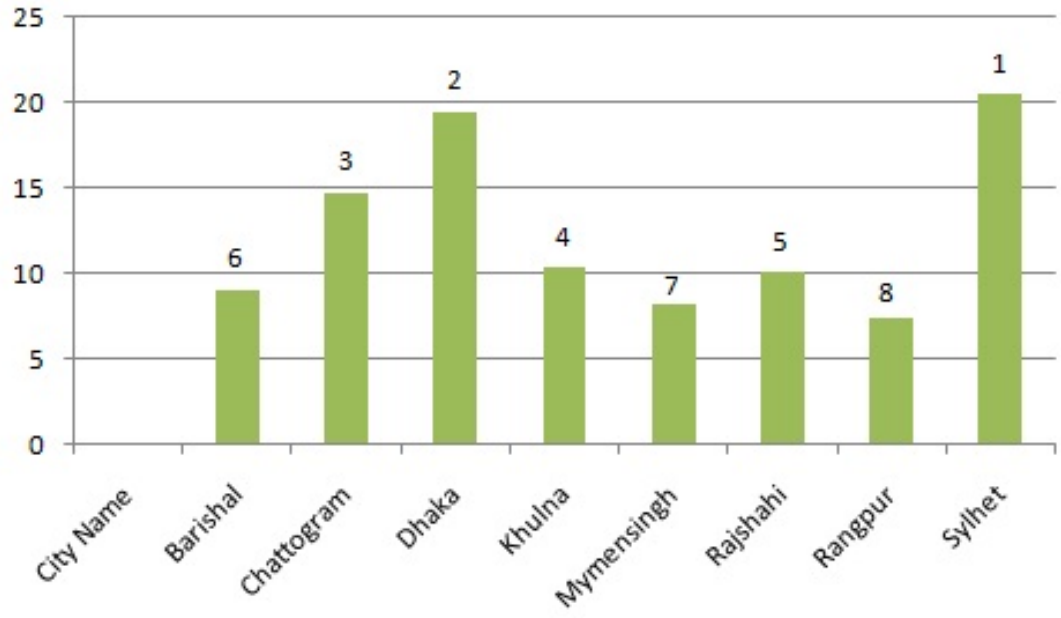



Figure 4.4: City rank in *DFA*.

### 4.5.4 Overall Rank

indexOverall Rank The overall rank has been computed over the final score of each city's in three dimensions. In summary results, we have found Barishal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet rank as 7, 2, 1, 4, 5, 6, 8 and 3 respectively. Here, Sylhet has got the 1st rank and Rangpur has got the 8th rank. The graph 4.5 expresses the overall ranking of eight divisional cities of Bangladesh.

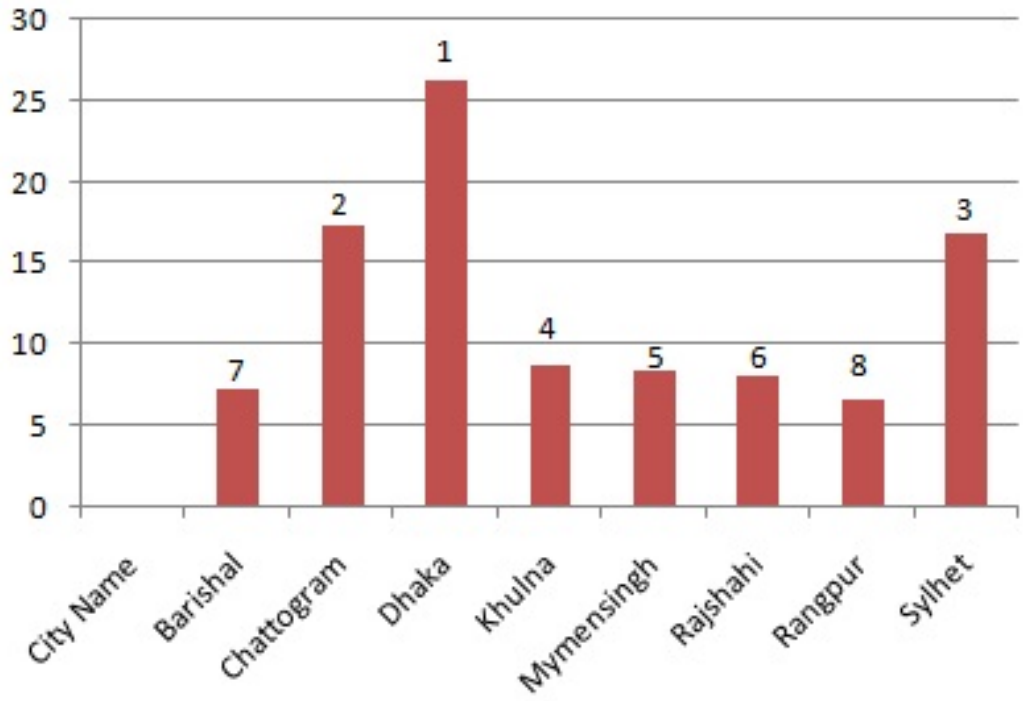



Figure 4.5: Overall ranking.

## 4.6 Result Comparison

Here, we compare our results with BBS's [Bangladesh Bureau of Statistics] data. Figure 4.6 shows the score comparison between BBS's and our computed data in *Economic Establishments* dimension. Here, we find some deviations for the following reasons:

- BBS results are derived from the conducted district wise surveys, on the other hand our SSM computes rank from the city wise data.
- SSM uses small amount of data while BBS uses big amount of data.
- In SSM, we use three dimensions but BBS uses only two dimensions.

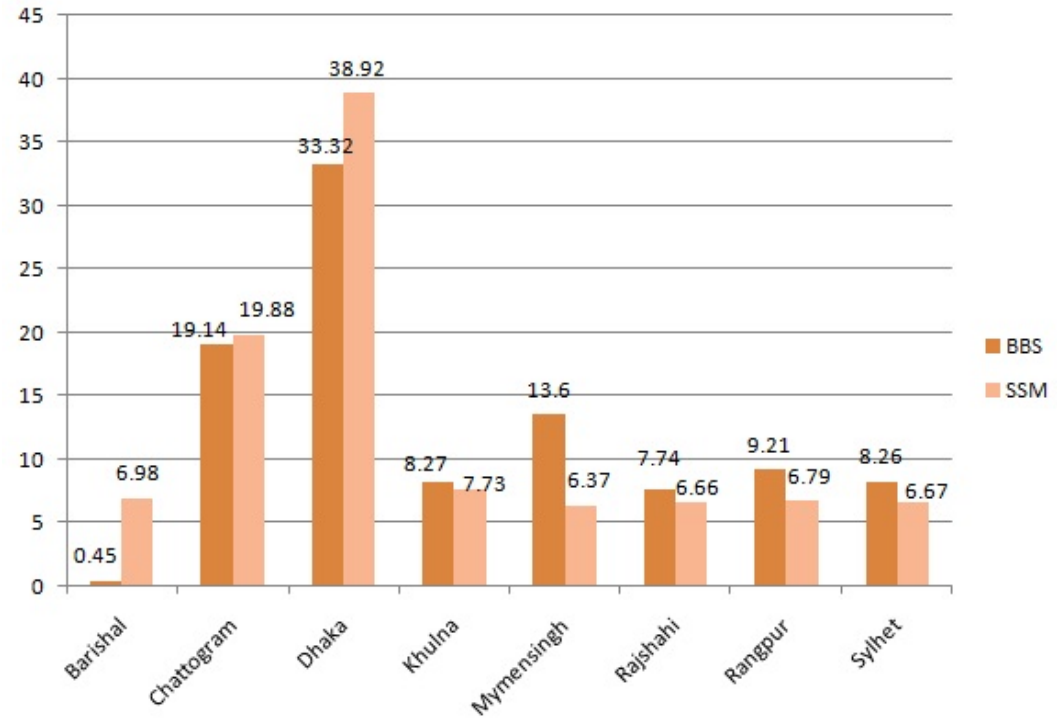



Figure 4.6: Comparison between scores computed by SSM and BBS for *EE* dimension.

Figure 4.7 shows the ranking obtained from BBS and our SSM for EE dimension. Here, we have found that four cities ranking match with BBS ranking (Dhaka, Chittagong, Rajshahi and Sylhet).

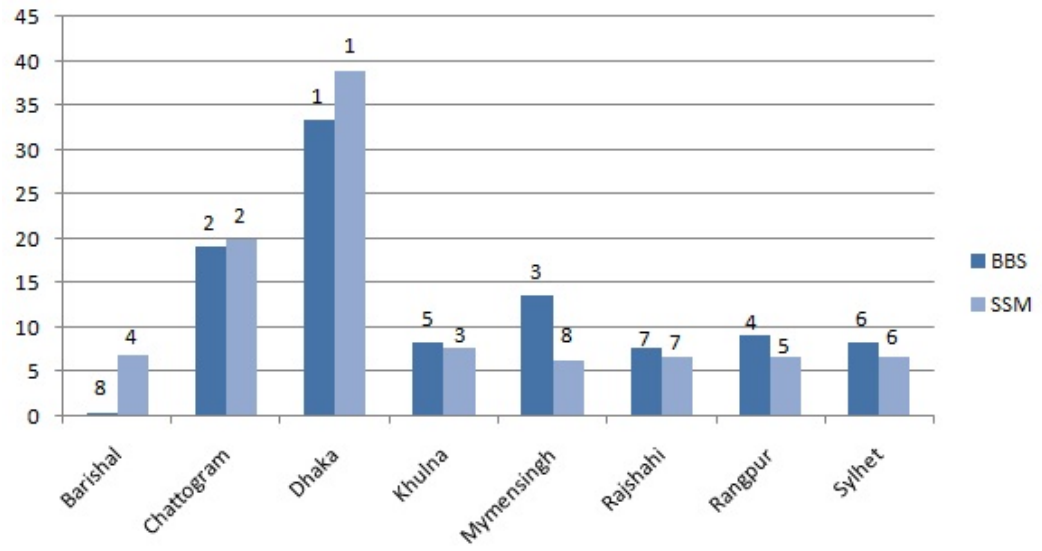



Figure 4.7: Comparison between ranking computed by SSM and BBS for *EE* dimension.

Figure 4.8 describes the comparison of scores computed by BBS and SSM for *Occupation* dimension. Here also, we find some variations in score because:

- SSM considers non-residents while in BBS collected data with only residents.
- BBS collected data occupation district wise. On the other hand, in SSM we count only city base data.

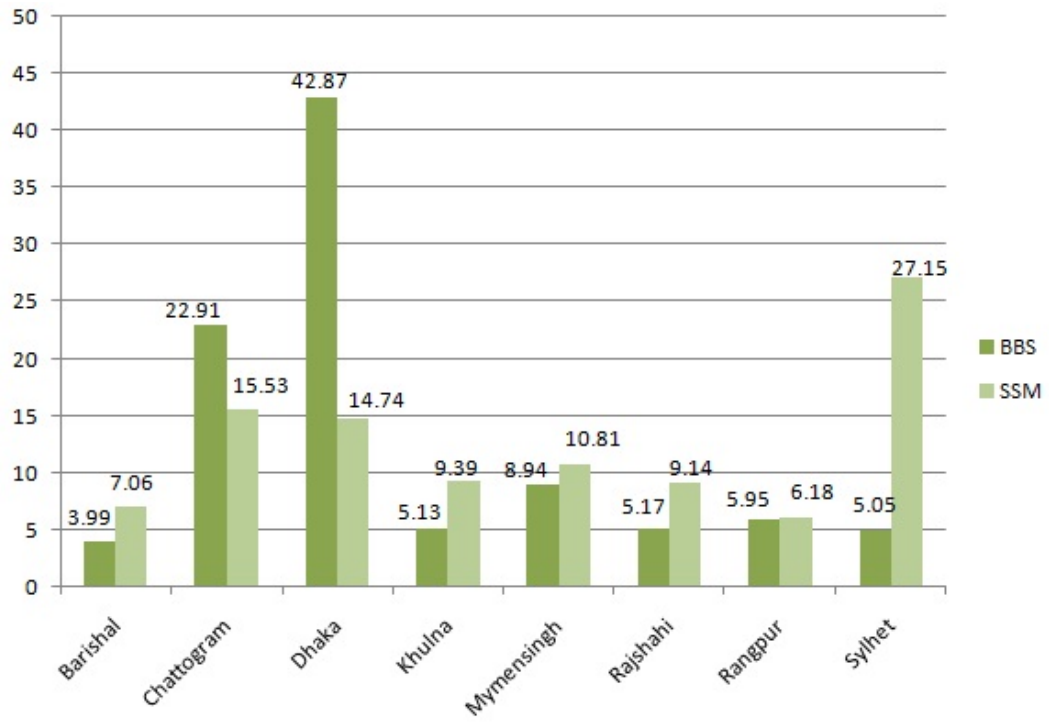



Figure 4.8: Comparison of scores between SSM and BBS for *Occupation*.

The figure 4.9 shows the comparison in ranking between BBS and SSM for occupation. Here we have found dissimilarities in cities's ranking and BBS's ranking (Barishal, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet). Only Chittagong matches.

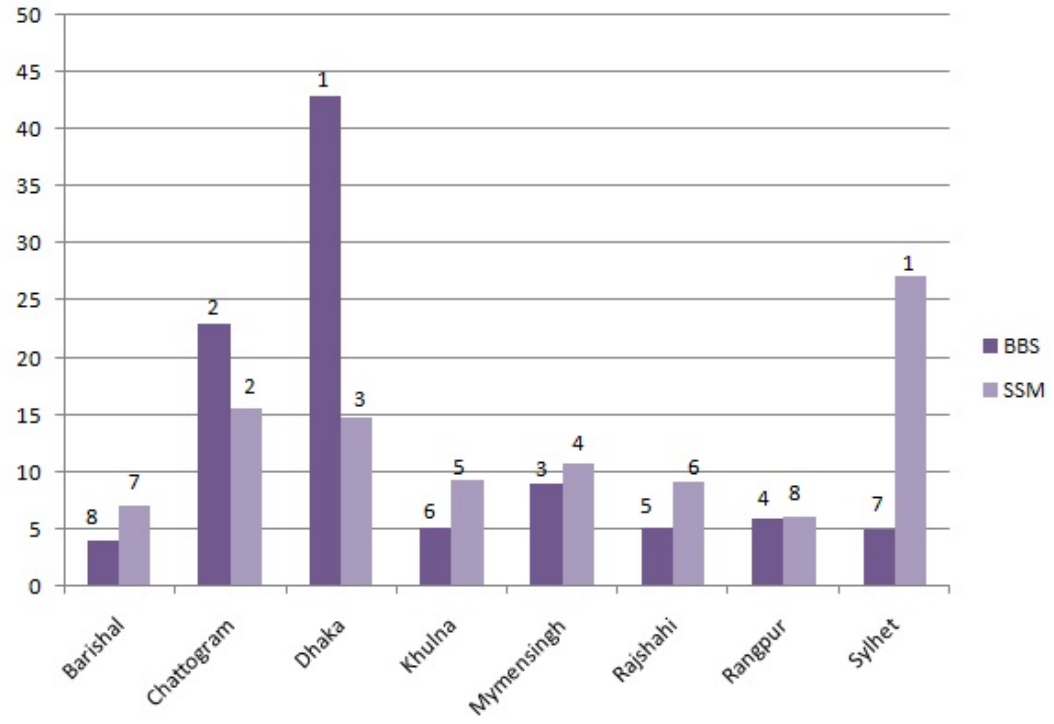



Figure 4.9: Comparison of ranking computed by SSM and BBS for *Occupation*.

Now, Figure 4.10 shows the comparison in overall score between BBS and SSM. Here, we find some differences in the score for the reasons described as follows.

- BBS results conducted district wise, on the other hand SSM uses city wise data.
- SSM uses small amount of data, while BBS uses big amount of data.
- SSM uses three dimensions, but BBS uses only two dimensions.
- SSM considers non-residents, while BBS considers only residents.

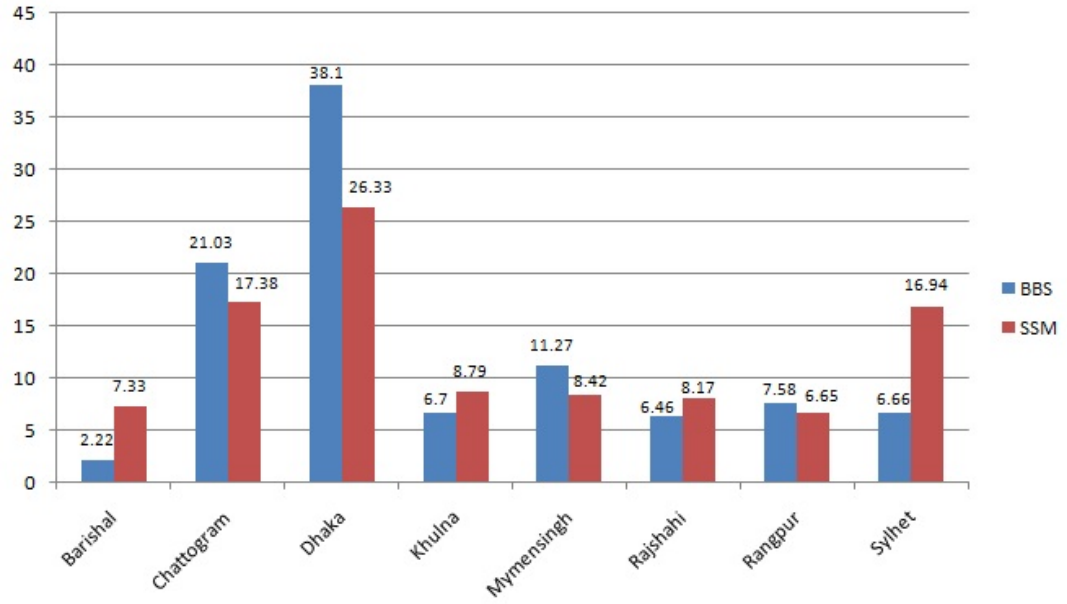



Figure 4.10: Comparison of overall score between SSM and BBS.

Figure 4.11 shows the comparison of overall ranking between BBS method and SSM. Only Dhaka and Chittagong match in overall ranks.

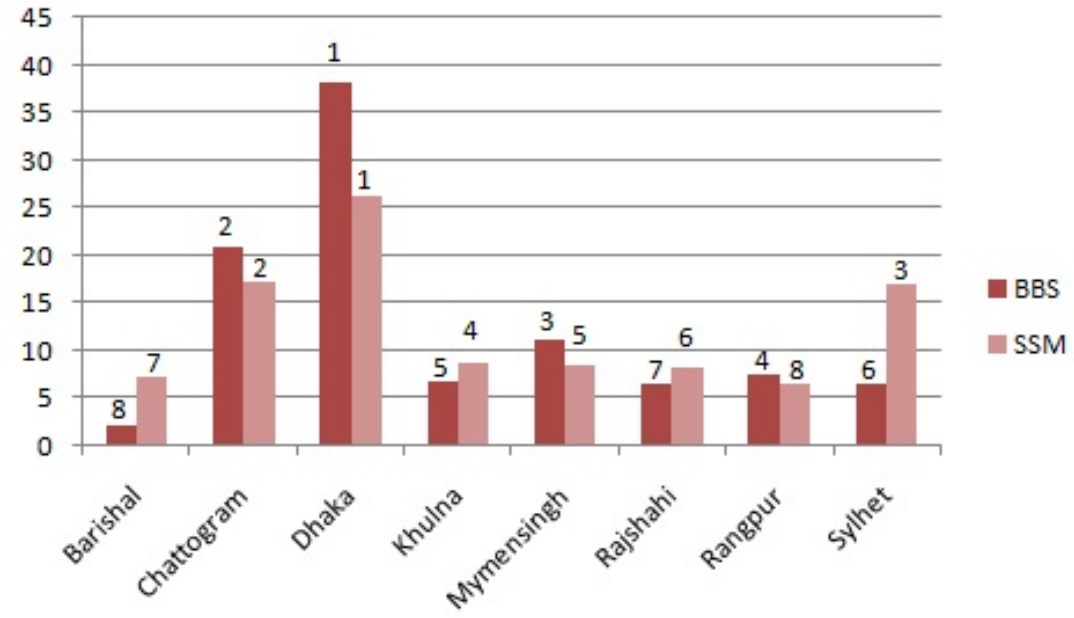



Figure 4.11: Comparison between overall ranking between SSM and BBS.

## 4.7 Limitations in Data Pre-Processing

The models we have built are not free of error. The probable causes for errors along with a complete analysis are given below:

1. **Language Problem:** Some users upload posts using English and Bengali both. Our model is designed for only English. For example, *"ami akta natun gari kinsi"*. Our framework is not able to extract financial data from these types of posts.

2. **One Count Problem:** Sometimes, our model counts only one activity while a user upload two or more financial activities in a post. Such as an example, *'Last week, I enjoyed a long trip, from Sunday to Wednesday i was in Cox's Bazar and from Thursday to Saturday I was in Sajek'*. From this example, our model counts here only one activity "trip" but in here two activities occurred one is travel in cox's bazaar and the another is travel in Sajek.

3. **VIP Keeps Out:** Some valuable people could not be automatically included in our research. Many business-VIP who have good business but there information are not available in social or web media. Our framework is based on only available data about the financial activities of general people.

## 4.8 Summary Results

Table 4.10 shows the summary results of our proposed framework.

| | **Overall** | | **EE** | | **Occ.** | | **DFA** | |
|---|---|---|---|---|---|---|---|---|
| | ***Rank*** | ***Score*** | ***Rank*** | ***Score*** | ***Rank*** | ***Score*** | ***Rank*** | ***Score*** |
| **Barishal** | 7 | 7.33 | 4 | 6.98 | 7 | 7.06 | 6 | 9.1 |
| **Chittagong** | 2 | 17.38 | 2 | 19.88 | 2 | 15.53 | 3 | 14.79 |
| **Dhaka** | 1 | 26.33 | 1 | 38.92 | 3 | 14.74 | 2 | 19.49 |
| **Khulna** | 4 | 8.79 | 3 | 7.73 | 5 | 9.39 | 4 | 10.35 |
| **Mymensingh** | 5 | 8.42 | 8 | 6.37 | 4 | 10.81 | 7 | 8.2 |
| **Rajshahi** | 6 | 8.17 | 7 | 6.66 | 6 | 9.14 | 5 | 10.08 |
| **Rangpur** | 8 | 6.65 | 5 | 6.79 | 8 | 6.18 | 8 | 7.49 |
| **Sylhet** | 3 | 16.94 | 6 | 6.67 | 1 | 27.15 | 1 | 20.5 |

Table 4.10: Summary results.

# Chapter 5

# Conclusions

Financial solvency measurement publishes a directory of national business statistics and overall national economy status. In most countries, the authorities carry out an economic census every five or six years. During a census, they gather lot of valuable information about financial institutions, employees, sales, receipts, assets etc. The data getting from the economic census is very important for industry, community and business because:

- Statistics from the economic census provide policy-makers with evidence-based information used to make appropriate programmatic decisions.
- Census data is used as the basis for a variety of economic activities such as gross domestic product (GDP), national income and commodity accounting (NIPA), and producer price index (PPI).
- To determine organizational structure and product trends the trade and business associations use economic census data for finding key business facts.
- To make decisions about capital investments, operating sites and product development, the individual businesses use the economic census data.

It is a difficult work process due to manual demographic data collection. The manual data collection process increases the costs lot of money and engaged huge people.

Therefore, we have proposed an automated framework called Solvency Sensing Method (SSM), in which data has been collected from social and web media. We measure solvency using Economic Establishment (EE), Occupation, and web based Daily Financial Activity (DFA). Our method saves money, time and effort.

In this research, we have found Dhaka division scored 1st rank, Chittagong scores 2nd, Sylhet 3rd, Khulna 4th, Mymensingh 5th, Rajshahi 6th, Barishal 7th and Rangpur 8th.

We would like to continue our research with a large number of social media users. If we get more data the results will be more accurate. We would like to include more financial indicators. Here, we ignored natural disasters, environmental and time impact. In future we would like to include those for better accuracy.

# References


[1] BBS, *Report on economic census 2013*. Ministry of Planning, Statistics and Information Division, Government of Bangladesh, Dhaka, Bangladesh, 2013.

[2] CSO, *Sixth Economic Census*. Government of India, Ministry of Statistics Program Implementation, National Statistical Commission, New Delhi, India, 2013.

[3] T. C. Federation, "Indicators of financial condition: A comparison of the city of chicago to 12 other u.s. cities," *in The Civic Federation, USA*, vol. 504, pp. 1–60, 2015.

[4] X. Wang, L. Dennis, and Y. S. Tu, "Measuring financial condition: A study of us states," *Public Budgeting & Finance*, vol. 27, no. 2, pp. 1–21, 2007.

[5] T. C. Federation, "Annual financial analysis 2017," *in The Civic Federation, USA*, vol. 709, pp. 1–58, 2017.

[6] B. Gadanecz and K. Jayaram, "Measures of financial stability-a review," *Irving Fisher Committee Bulletin*, vol. 31, pp. 365–383, 2008.

[7] R. Steinbach, *Growth in Low-Income Countries: Evolution, Prospects, and Policies*. The World Bank, 2019.

[8] K. Schwab and X. Sala-i Martín, *The global competitiveness report 2015–2016, Geneva*. The World Economic Forum, 2015.

[9] S. Alkire, A. Conconi, M. Pinilla-Roncancio, and A. Vaz, *How to Build a National Multidimensional Poverty Index (MPI): Using the MPI to inform the SDGs*. United Nations Development Programme (UNDP), 2018.

[10] W. Bank, *Global Economic Prospects, June 2019: Heightened Tensions, Subdued Investment*. The World Bank, 2019.

[11] L. Auria and R. A. Moro, "Support vector machines (svm) as a technique for solvency analysis," *DIW Berlin discussion paper, Berlin*, no. 811, pp. 1–19, 2008.

[12] T. H. Silva, P. O. V. de Melo, J. M. Almeida, M. Musolesi, and A. A. Loureiro, "You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare," in *Eighth International AAAI Conference on Weblogs and Social Media*, pp. 466–475, AAAI, 2014.

[13] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *2013 IEEE 13th International Conference on Data Mining*, pp. 1151–1156, IEEE, 2013.

[14] G.-J. Qi, C. C. Aggarwal, and T. Huang, "Community detection with edge content in social media networks," in *2012 IEEE 28th International Conference on Data Engineering*, pp. 534–545, IEEE, 2012.

[15] T. Hoang, P. H. Cher, P. K. Prasetyo, and E.-P. Lim, "Crowdsensing and analyzing micro-event tweets for public transportation insights," in *2016 IEEE International Conference on Big Data (Big Data)*, pp. 2157–2166, IEEE, 2016.

[16] UNSD, *International Standard Industrial Classification of All Economic Activities (ISIC) Revision 4*. New York: United Nations Statistics Division, 2007.

[17] L. V. Subramaniam, S. Roy, T. A. Faruquie, and S. Negi, "A survey of types of text noise and techniques to handle noisy text," in *Proceedings of The Third Workshop on Analytics for Noisy Unstructured Text Data*, pp. 115–122, ACM, 2009.

[18] C. Sengstock and M. Gertz, "Latent geographic feature extraction from social media," in *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, pp. 149–158, ACM, 2012.

[19] H. Li, H. Ji, and L. Zhao, "Social event extraction: Task, challenges and techniques," in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 526–532, IEEE, 2015.

[20] J. Tang, Y. Chang, and H. Liu, "Mining social media with social theories: a survey," *ACM Sigkdd Explorations Newsletter*, vol. 15, no. 2, pp. 20–29, 2014.

[21] I. Guellil and K. Boukhalfa, "Social big data mining: A survey focused on opinion mining and sentiments analysis," in *2015 12th International Symposium on Programming and Systems (ISPS)*, pp. 1–10, IEEE, 2015.

[22] R. Rabade, N. Mishra, and S. Sharma, "Survey of influential user identification techniques in online social networks," in *Recent advances in intelligent informatics*, pp. 359–370, Springer, 2014.

[23] A. Culotta, "Estimating county health statistics with twitter," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pp. 1335–1344, ACM, 2014.

[24] M. E. Ali and M. T. Islam, *A Temporal Psycholinguistics Approach to Identity Resolution of Social Media Users*. Bangladesh University of Engineering and Technology, 2017.

# Index

# Appendix A

# Algorithms

## A.1 Ranking

**Algorithm 1** Rank Calculation

**Input:** $DATA[][], ROW, COL, W_i[ROW][COL]$
**Output:** $RANK$ (Rank values in the array)
0: $Set\ I \leftarrow 0$, $J \leftarrow 0$, $RANK[]$, $SCORE[]$
0: $SCORE \leftarrow\ SCORE - CALCULATION(DATA[][], W_i[ROW][COL], ROW, COL)$
{Rank calculation}
0: **for** $I \leftarrow 1$ to $COL$ **do**
0: $RANK[I] \leftarrow rank(SCORE[I])$
0: **end for**
{Print Score}
0: **for** $I \leftarrow 1$ to $COL$ **do**
0: $Print(SCORE[I])$
0: **end for**
{Print Rank}
0: **for** $I \leftarrow 1$ to $COL$ **do**
0: $Print(RANK[I])$
0: **end for**
=0

## A.2 Scoring

**Algorithm 2** Score Calculation

```
0: function SCORE-CALCULATION(DATA[][], W_i[][], ROW, COL)
0:     Set I ← 0, J ← 0, Sum ← 0, addrow[], avg[][], aw[][], score[]
  {Adding Row}
0:     for I ← 1 to ROW do
0:         for J ← 1 to COL do
0:             Sum ← Sum + DATA[I][J]
0:         end for
0:         addrow[I] ← Sum
0:         Sum ← 0
0:     end for
  {Row level averaging}
0:     for I ← 1 to ROW do
0:         for J ← 1 to COL do
0:             avg[I][J] ← ((DATA[I][J]/addrow[I]) * 100)
0:         end for
0:     end for
  {Weighted average without percentage}
0:     for I ← 1 to ROW do
0:         for J ← 1 to COL do
0:             aw[I][J] ← (avg[I][J] * W_i[I])
0:         end for
0:     end for
  {column wise summation and score calculation}
0:     for I ← 1 to COL do
0:         for J ← 1 to ROW do
0:             Sum ← (Sum + aw[I][J])
0:         end for
0:         score[J] ← Sum
0:         Sum ← 0
0:     end for
0:     return score
0: end function
  =0
```

# Appendix B

# Categorical Data

## B.1 Agriculture, Forestry and Fishing

Table B.1: Agriculture, Forestry and Fishing

| Indicator-1 | Description |
|---|---|
| AFF | **Crop and animal production, hunting and related service activities** |
| | Growing of non-perennial crops |
| | Growing of cereals (except rice), leguminous crops and oil seeds |
| | Growing of rice |
| | Growing of vegetables and melons, roots and tubers |
| | Growing of sugar cane |
| | Growing of tobacco |
| | Growing of fibre crops |
| | Growing of other non-perennial crops |
| | Growing of perennial crops |
| | Growing of grapes |
| | Growing of tropical and subtropical fruits |
| | Growing of citrus fruits |
| | Growing of pome fruits and stone fruits |
| | Growing of other tree and bush fruits and nuts |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Growing of oleaginous fruits |
| | Growing of beverage crops |
| | Growing of spices, aromatic, drug and pharmaceutical crops |
| | Growing of other perennial crops |
| | Plant propagation |
| | Animal production |
| | Raising of cattle and buffaloes |
| | Raising of horses and other equines |
| | Raising of camels and camelids |
| | Raising of sheep and goats |
| | Raising of swine/pigs |
| | Raising of poultry |
| | Raising of other animals |
| | Mixed farming |
| | Support activities to agriculture and post-harvest crop activities |
| | Support activities for crop production |
| | Support activities for animal production |
| | Post-harvest crop activities |
| | Seed processing for propagation |
| | Hunting, trapping and related service activities |
| | **Forestry and logging** |
| | Silviculture and other forestry activities |
| | Logging |
| | Gathering of non-wood forest products |
| | Support services to forestry |
| | **Fishing and aquaculture** |
| | Fishing |
| | Marine fishing |
| | Freshwater fishing |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Aquaculture |
| | Marine aquaculture |
| | Freshwater aquaculture |

## B.2 Mining and Quarrying

Table B.2: Mining and Quarrying.

| Indicator-2 | Description |
|---|---|
| MAQ | **Mining of coal and lignite** |
| | Mining of hard coal |
| | Mining of lignite |
| | **Extraction of crude petroleum and natural gas** |
| | Extraction of crude petroleum |
| | Extraction of natural gas |
| | **Mining of metal ores** |
| | Mining of iron ores |
| | Mining of non-ferrous metal ores |
| | Mining of uranium and thorium ores |
| | Mining of other non-ferrous metal ores |
| | **Other mining and quarrying** |
| | Quarrying of stone, sand and clay |
| | Mining and quarrying n.e.c. |
| | Mining of chemical and fertilizer minerals |
| | Extraction of peat |
| | Extraction of salt |
| | Other mining and quarrying |
| | **Mining support service activities** |
| | Support activities for petroleum and natural gas extraction |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Support activities for other mining and quarrying |

## B.3 Manufacturing

Table B.3: Manufacturing

| **Indicator-3** | **Description** |
|---|---|
| MFG | **Manufacture of food products** |
| | Processing and preserving of meat |
| | Processing and preserving of fish, crustaceans and molluscs |
| | Processing and preserving of fruit and vegetables |
| | Manufacture of vegetable and animal oils and fats |
| | Manufacture of dairy products |
| | Manufacture of grain mill products, starches and starch products |
| | Manufacture of grain mill products |
| | Manufacture of starches and starch products |
| | Manufacture of other food products |
| | Manufacture of bakery products |
| | Manufacture of sugar |
| | Manufacture of cocoa, chocolate and sugar confectionery |
| | Manufacture of macaroni, noodles, couscous and similar farinaceous products |
| | Manufacture of prepared meals and dishes |
| | Manufacture of other food products n.e.c. |
| | Manufacture of prepared animal feeds |
| | **Manufacture of beverages** |
| | Distilling, rectifying and blending of spirits |
| | Manufacture of wines |
| | Manufacture of malt liquors and malt |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Manufacture of soft drinks; production of mineral waters and other bottled waters |
| | **Manufacture of tobacco products** |
| | Manufacture of tobacco products |
| | **Manufacture of textiles** |
| | Spinning, weaving and finishing of textiles |
| | Preparation and spinning of textile fibres |
| | Weaving of textiles |
| | Finishing of textiles |
| | Manufacture of other textiles |
| | Manufacture of knitted and crocheted fabrics |
| | Manufacture of made-up textile articles, except apparel |
| | Manufacture of carpets and rugs |
| | Manufacture of cordage, rope, twine and netting |
| | Manufacture of other textiles n.e.c. |
| | **Manufacture of wearing apparel** |
| | Manufacture of wearing apparel, except fur apparel |
| | Manufacture of articles of fur |
| | Manufacture of knitted and crocheted apparel |
| | **Manufacture of leather and related products** |
| | Tanning and dressing of leather; manufacture of luggage, handbags, saddlery and harness; dressing and dyeing of fur |
| | Tanning and dressing of leather; dressing and dyeing of fur |
| | Manufacture of luggage, handbags and the like, saddlery and harness |
| | Manufacture of footwear |
| | **Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials** |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Sawmilling and planing of wood |
| | Manufacture of products of wood, cork, straw and plaiting materials |
| | Manufacture of veneer sheets and wood-based panels |
| | Manufacture of builders' carpentry and joinery |
| | Manufacture of wooden containers |
| | Manufacture of other products of wood; manufacture of articles of cork, straw and plaiting materials |
| | **Manufacture of paper and paper products** |
| | Manufacture of pulp, paper and paperboard |
| | Manufacture of corrugated paper and paperboard and of containers of paper and paperboard |
| | Manufacture of other articles of paper and paperboard |
| | **Printing and reproduction of recorded media** |
| | Printing and service activities related to printing |
| | Printing |
| | Service activities related to printing |
| | Reproduction of recorded media |
| | **Manufacture of coke and refined petroleum products** |
| | Manufacture of coke oven products |
| | Manufacture of refined petroleum products |
| | **Manufacture of chemicals and chemical products** |
| | Manufacture of basic chemicals, fertilizers and nitrogen compounds, plastics and synthetic rubber in primary forms |
| | Manufacture of basic chemicals |
| | Manufacture of fertilizers and nitrogen compounds |
| | Manufacture of plastics and synthetic rubber in primary forms |
| | Manufacture of other chemical products |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Manufacture of pesticides and other agrochemical products |
| | Manufacture of paints, varnishes and similar coatings, printing ink and mastics |
| | Manufacture of soap and detergents, cleaning and polishing preparations, perfumes and toilet preparations |
| | Manufacture of other chemical products n.e.c. |
| | Manufacture of man-made fibres |
| | **Manufacture of pharmaceuticals, medicinal chemical and botanical products** |
| | Manufacture of pharmaceuticals, medicinal chemical and botanical products |
| | **Manufacture of rubber and plastics products** |
| | Manufacture of rubber products |
| | Manufacture of rubber tyres and tubes; retreading and rebuilding of rubber tyres |
| | Manufacture of other rubber products |
| | Manufacture of plastics products |
| | **Manufacture of other non-metallic mineral products** |
| | Manufacture of glass and glass products |
| | Manufacture of non-metallic mineral products n.e.c. |
| | Manufacture of refractory products |
| | Manufacture of clay building materials |
| | Manufacture of other porcelain and ceramic products |
| | Manufacture of cement, lime and plaster |
| | Manufacture of articles of concrete, cement and plaster |
| | Cutting, shaping and finishing of stone |
| | Manufacture of other non-metallic mineral products n.e.c. |
| | **Manufacture of basic metals** |
| | Manufacture of basic iron and steel |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Manufacture of basic precious and other non-ferrous metals |
| | Casting of metals |
| | Casting of iron and steel |
| | Casting of non-ferrous metals |
| | **Manufacture of fabricated metal products, except machinery and equipment** |
| | Manufacture of structural metal products, tanks, reservoirs and steam generators |
| | Manufacture of structural metal products |
| | Manufacture of tanks, reservoirs and containers of metal |
| | Manufacture of steam generators, except central heating hot water boilers |
| | Manufacture of weapons and ammunition |
| | Manufacture of other fabricated metal products; metal-working service activities |
| | Forging, pressing, stamping and roll-forming of metal; powder metallurgy |
| | Treatment and coating of metals; machining |
| | Manufacture of cutlery, hand tools and general hardware |
| | Manufacture of other fabricated metal products n.e.c. |
| | **Manufacture of computer, electronic and optical products** |
| | Manufacture of electronic components and boards |
| | Manufacture of computers and peripheral equipment |
| | Manufacture of communication equipment |
| | Manufacture of consumer electronics |
| | Manufacture of measuring, testing, navigating and control equipment; watches and clocks |
| | Manufacture of measuring, testing, navigating and control equipment |
| | Manufacture of watches and clocks |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Manufacture of irradiation, electromedical and electrotherapeutic equipment |
| | Manufacture of optical instruments and photographic equipment |
| | Manufacture of magnetic and optical media |
| | **Manufacture of electrical equipment** |
| | Manufacture of electric motors, generators, transformers and electricity distribution and control apparatus |
| | Manufacture of batteries and accumulators |
| | Manufacture of wiring and wiring devices |
| | Manufacture of fibre optic cables |
| | Manufacture of other electronic and electric wires and cables |
| | Manufacture of wiring devices |
| | Manufacture of electric lighting equipment |
| | Manufacture of domestic appliances |
| | Manufacture of other electrical equipment |
| | **Manufacture of machinery and equipment n.e.c.** |
| | Manufacture of general-purpose machinery |
| | Manufacture of engines and turbines, except aircraft, vehicle and cycle engines |
| | Manufacture of fluid power equipment |
| | Manufacture of other pumps, compressors, taps and valves |
| | Manufacture of bearings, gears, gearing and driving elements |
| | Manufacture of ovens, furnaces and furnace burners |
| | Manufacture of lifting and handling equipment |
| | Manufacture of office machinery and equipment (except computers and peripheral equipment) |
| | Manufacture of power-driven hand tools |
| | Manufacture of other general-purpose machinery |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Manufacture of special-purpose machinery |
| | Manufacture of agricultural and forestry machinery |
| | Manufacture of metal-forming machinery and machine tools |
| | Manufacture of machinery for metallurgy |
| | Manufacture of machinery for mining, quarrying and construction |
| | Manufacture of machinery for food, beverage and tobacco processing |
| | Manufacture of machinery for textile, apparel and leather production |
| | Manufacture of other special-purpose machinery |
| | **Manufacture of motor vehicles, trailers and semi-trailers** |
| | Manufacture of motor vehicles |
| | Manufacture of bodies (coachwork) for motor vehicles; manufacture of trailers and semi-trailers |
| | Manufacture of parts and accessories for motor vehicles |
| | **Manufacture of other transport equipment** |
| | Building of ships and boats |
| | Building of ships and floating structures |
| | Building of pleasure and sporting boats |
| | Manufacture of railway locomotives and rolling stock |
| | Manufacture of air and spacecraft and related machinery |
| | Manufacture of military fighting vehicles |
| | Manufacture of transport equipment n.e.c. |
| | Manufacture of motorcycles |
| | Manufacture of bicycles and invalid carriages |
| | Manufacture of other transport equipment n.e.c. |
| | **Manufacture of furniture** |
| | Manufacture of furniture |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | **Other manufacturing** |
| | Manufacture of jewellery, bijouterie and related articles |
| | Manufacture of jewellery and related articles |
| | Manufacture of imitation jewellery and related articles |
| | Manufacture of musical instruments |
| | Manufacture of sports goods |
| | Manufacture of games and toys |
| | Manufacture of medical and dental instruments and supplies |
| | Other manufacturing n.e.c. |
| | **Repair and installation of machinery and equipment** |
| | Repair of fabricated metal products, machinery and equipment |
| | Repair of fabricated metal products |
| | Repair of machinery |
| | Repair of electronic and optical equipment |
| | Repair of electrical equipment |
| | Repair of transport equipment, except motor vehicles |
| | Repair of other equipment |
| | Installation of industrial machinery and equipment |

## B.4 Wholesale and Retail Trade

Table B.4: Wholesales and Retail Trades, Repair Industries.

| Indicator-7 | Description |
|---|---|
| WRT | **Wholesale and retail trade and repair of motor vehicles and motorcycles** |
| | Sale of motor vehicles |
| | Maintenance and repair of motor vehicles |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Sale of motor vehicle parts and accessories |
| | Sale, maintenance and repair of motorcycles and related parts and accessories |
| | **Wholesale trade, except of motor vehicles and motorcycles** |
| | Wholesale on a fee or contract basis |
| | Wholesale of agricultural raw materials and live animals |
| | Wholesale of food, beverages and tobacco |
| | Wholesale of household goods |
| | Wholesale of textiles, clothing and footwear |
| | Wholesale of other household goods |
| | Wholesale of machinery, equipment and supplies |
| | Wholesale of computers, computer peripheral equipment and software |
| | Wholesale of electronic and telecommunications equipment and parts |
| | Wholesale of agricultural machinery, equipment and supplies |
| | Wholesale of other machinery and equipment |
| | Other specialized wholesale |
| | Wholesale of solid, liquid and gaseous fuels and related products |
| | Wholesale of metals and metal ores |
| | Wholesale of construction materials, hardware, plumbing and heating equipment and supplies |
| | Wholesale of waste and scrap and other products n.e.c. |
| | Non-specialized wholesale trade |
| | **Retail trade, except of motor vehicles and motorcycles** |
| | Retail sale in non-specialized stores |
| | Retail sale in non-specialized stores with food, beverages or tobacco predominating |
| | Other retail sale in non-specialized stores |
| | Retail sale of food, beverages and tobacco in specialized stores |
| | Retail sale of food in specialized stores |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Retail sale of beverages in specialized stores |
| | Retail sale of tobacco products in specialized stores |
| | Retail sale of automotive fuel in specialized stores |
| | Retail sale of information and communications equipment in specialized stores |
| | Retail sale of computers, peripheral units, software and telecommunications equipment in specialized stores |
| | Retail sale of audio and video equipment in specialized stores |
| | Retail sale of other household equipment in specialized stores |
| | Retail sale of textiles in specialized stores |
| | Retail sale of hardware, paints and glass in specialized stores |
| | Retail sale of carpets, rugs, wall and floor coverings in specialized stores |
| | Retail sale of electrical household appliances, furniture, lighting equipment and other household articles in specialized stores |
| | Retail sale of cultural and recreation goods in specialized stores |
| | Retail sale of books, newspapers and stationary in specialized stores |
| | Retail sale of music and video recordings in specialized stores |
| | Retail sale of sporting equipment in specialized stores |
| | Retail sale of games and toys in specialized stores |
| | Retail sale of other goods in specialized stores |
| | Retail sale of clothing, footwear and leather articles in specialized stores |
| | Retail sale of pharmaceutical and medical goods, cosmetic and toilet articles in specialized stores |
| | Other retail sale of new goods in specialized stores |
| | Retail sale of second-hand goods |
| | Retail sale via stalls and markets |
| | Retail sale via stalls and markets of food, beverages and tobacco products |
| | Retail sale via stalls and markets of textiles, clothing and footwear |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Retail sale via stalls and markets of other goods |
| | Retail trade not in stores, stalls or markets |
| | Retail sale via mail order houses or via Internet |
| | Other retail sale not in stores, stalls or markets |

## B.5 Transportation and Storage

Table B.5: Transportation and Storage.

| Indicator-8 | Description |
|---|---|
| TAS | **Land transport and transport via pipelines** |
| | Transport via railways |
| | Passenger rail transport, interurban |
| | Freight rail transport |
| | Other land transport |
| | Urban and suburban passenger land transport |
| | Other passenger land transport |
| | Freight transport by road |
| | Transport via pipeline |
| | **Water transport** |
| | Sea and coastal water transport |
| | Sea and coastal passenger water transport |
| | Sea and coastal freight water transport |
| | Inland water transport |
| | Inland passenger water transport |
| | Inland freight water transport |
| | **Air transport** |
| | Passenger air transport |
| | Freight air transport |
| | **Warehousing and support activities for transportation** |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Warehousing and storage |
| | Support activities for transportation |
| | Service activities incidental to land transportation |
| | Service activities incidental to water transportation |
| | Service activities incidental to air transportation |
| | Cargo handling |
| | Other transportation support activities |
| | **Postal and courier activities** |
| | Postal activities |
| | Courier activities |

## B.6 Information Technologies and Communication

Table B.6: Information Technologies and Communication.

| Indicator-10 | Description |
|---|---|
| IC | **Publishing activities** |
| | Publishing of books, periodicals and other publishing activities |
| | Book publishing |
| | Publishing of directories and mailing lists |
| | Publishing of newspapers, journals and periodicals |
| | Other publishing activities |
| | Software publishing |
| | **Motion picture, video and television programme production, sound recording and music publishing activities** |
| | Motion picture, video and television programme activities |
| | Motion picture, video and television programme production activities |
| | Motion picture, video and television programme post-production activities |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Motion picture, video and television programme distribution activities |
| | Motion picture projection activities |
| | Sound recording and music publishing activities |
| | **Programming and broadcasting activities** |
| | Radio broadcasting |
| | Television programming and broadcasting activities |
| | **Telecommunications** |
| | Wired telecommunications activities |
| | Wireless telecommunications activities |
| | Satellite telecommunications activities |
| | Other telecommunications activities |
| | **Computer programming, consultancy and related activities** |
| | Computer programming activities |
| | Computer consultancy and computer facilities management activities |
| | Other information technology and computer service activities |
| | **Information service activities** |
| | Data processing, hosting and related activities; web portals |
| | Data processing, hosting and related activities |
| | Web portals |
| | Other information service activities |
| | News agency activities |
| | Other information service activities |

## B.7 Financial and Insurance Activities

Table B.7: Financial and Insurance Activities.

| Indicator-11 | Description |
|---|---|

*continued on next page*

*continued from previous page*

| | |
|---|---|
| FIA | **Financial service activities, except insurance and pension funding** |
| | Monetary intermediation |
| | Central banking |
| | Other monetary intermediation |
| | Activities of holding companies |
| | Trusts, funds and similar financial entities |
| | Other financial service activities, except insurance and pension funding activities |
| | Financial leasing |
| | Other credit granting |
| | Other financial service activities, except insurance and pension funding activities, n.e.c. |
| | **Insurance, reinsurance and pension funding, except compulsory social security** |
| | Insurance |
| | Life insurance |
| | Non-life insurance |
| | Reinsurance |
| | Pension funding |
| | **Activities auxiliary to financial service and insurance activities** |
| | Activities auxiliary to financial service activities, except insurance and pension funding |
| | Administration of financial markets |
| | Security and commodity contracts brokerage |
| | Other activities auxiliary to financial service activities |
| | Activities auxiliary to insurance and pension funding |
| | Risk and damage evaluation |
| | Activities of insurance agents and brokers |
| | Other activities auxiliary to insurance and pension funding |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Fund management activities |

## B.8 Professional, Scientific and Technical Activities

Table B.8: Professional, Scientific, and Technical Activity Data

| Indicator-13 | Description |
|---|---|
| PSTA | **Legal and accounting activities** |
| | Legal activities |
| | Accounting, bookkeeping and auditing activities; tax consultancy |
| | **Activities of head offices; management consultancy activities** |
| | Activities of head offices |
| | Management consultancy activities |
| | **Architectural and engineering activities; technical testing and analysis** |
| | Architectural and engineering activities and related technical consultancy |
| | Technical testing and analysis |
| | **Scientific research and development** |
| | Research and experimental development on natural sciences and engineering |
| | Research and experimental development on social sciences and humanities |
| | **Advertising and market research** |
| | Advertising |
| | Market research and public opinion polling |
| | **Other professional, scientific and technical activities** |
| | Specialized design activities |
| | Photographic activities |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Other professional, scientific and technical activities n.e.c. |
| | **Veterinary activities** |
| | Veterinary activities |

## B.9 Administrative and Support Services

Table B.9: Administrative and Support Services

| Indicator-14 | Description |
|---|---|
| ASSA | **Rental and leasing activities** |
| | Renting and leasing of motor vehicles |
| | Renting and leasing of personal and household goods |
| | Renting and leasing of recreational and sports goods |
| | Renting of video tapes and disks |
| | Renting and leasing of other personal and household goods |
| | Renting and leasing of other machinery, equipment and tangible goods |
| | Leasing of intellectual property and similar products, except copyrighted works |
| | **Employment activities** |
| | Activities of employment placement agencies |
| | Temporary employment agency activities |
| | Other human resources provision |
| | **Travel agency, tour operator, reservation service and related activities** |
| | Travel agency and tour operator activities |
| | Travel agency activities |
| | Tour operator activities |
| | Other reservation service and related activities |
| | **Security and investigation activities** |
| | Private security activities |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Security systems service activities |
| | Investigation activities |
| | **Services to buildings and landscape activities** |
| | Combined facilities support activities |
| | Cleaning activities |
| | General cleaning of buildings |
| | Other building and industrial cleaning activities |
| | Landscape care and maintenance service activities |
| | **Office administrative, office support and other business support activities** |
| | Office administrative and support activities |
| | Combined office administrative service activities |
| | Photocopying, document preparation and other specialized office support activities |
| | Activities of call centres |
| | Organization of conventions and trade shows |
| | Business support service activities n.e.c. |
| | Activities of collection agencies and credit bureaus |
| | Packaging activities |
| | Other business support service activities |

## B.10 Other Activities

Table B.10: Other Service Activities.

| Indicator-21 | Description |
|---|---|
| OSA | **Activities of membership organizations** |
| | Activities of business, employers and professional membership organizations |
| | Activities of business and employers membership organizations |
| | Activities of professional membership organizations |

*continued on next page*

*continued from previous page*

| | |
|---|---|
| | Activities of trade unions |
| | Activities of other membership organizations |
| | Activities of religious organizations |
| | Activities of political organizations |
| | Activities of other membership organizations n.e.c. |
| | **Repair of computers and personal and household goods** |
| | Repair of computers and communication equipment |
| | Repair of computers and peripheral equipment |
| | Repair of communication equipment |
| | Repair of personal and household goods |
| | Repair of consumer electronics |
| | Repair of household appliances and home and garden equipment |
| | Repair of footwear and leather goods |
| | Repair of furniture and home furnishings |
| | Repair of other personal and household goods |
| | **Other personal service activities** |
| | Washing and (dry-) cleaning of textile and fur products |
| | Hairdressing and other beauty treatment |
| | Funeral and related activities |
| | Other personal service activities. |

# Appendix C

# Data and Code

## C.1 Data for Economic Establishments

```
1 library('rvest')
2 url <- 'http://bangladeshyellowpages.com/search/
3 results.html?what=&where=sylhet'
4 webpage <- read_html(url)
5 node<- html_nodes(webpage ,'.search-result')
6 data <- html_text(node)
7
8 fileConn<-file("cc.csv")
9 eedata <- writeLines(data, "cc.csv")
10 close(fileConn)
```

## C.2 Social Media Data for Occupation and DFA

```
install.packages("Rfacebook")
install.packages("RCurl")
install.packages("httpuv")
install.packages("httr")
install.packages("rjson")
install.packages("tm")
install.packages("wordcloud")
install.packages("textclean")


library("Rfacebook")
library("RCurl")
library("httpuv")
library("httr")
library("rjson")
library("tm")
library("wordcloud")
library("textclean")

access_token <-
EAAIS2lAOiRABAAZADvqvOQysQDTHQShcCp0uQ4awpzlzQ5MSBj1rxub9E4li9RwWUMbp
OND6taK2aLqnHBrKtxZCkzUSD2kxYTBm0pOhpOuuGunwJZBl8k2149dhc2WaQAoCnxZBt9Y
TZCHQyT1uCCzBIWuG5f9cgeXmTuUZBO1fEZBziBbrWhA3bLUB2SqG7oZD

Options(RCurloptions =
list(verbose=FALSE,capath=system.file("CurlSLL","cacert.pem", package =
 "RCurl"),ssl.verifypeer=FALSE))
me <- getUsers("me", token= access_token)
myFriends <- getFriends(access_token,simplify = FALSE)
fileConn<-file("cc.csv")
fb_data <- writeLines(myFriend, "cc.csv")
close(fileConn)
```

## C.3 User Selection from Social Media Data

```
file1<-file("user.csv")
file2<-file("area.txt")
x<- grep("hometown", readLines("area.txt"), value = TRUE)
if(x == 0){
file.remove("user.csv")
}
close(file1)
close(file2)
```

## C.4 Pre-processing of EE Data

```
fileConn<-file("dd.csv")
xx <- readlines("dd.csv")
xx<-tm_map(xx,removeNumbers, lazy=TRUE, 'mc.cores=1')
xx<-tm_map(xx,stripWhitespace, lazy=TRUE, 'mc.cores=1')
xx<-tm_map(xx,removePunctuation, lazy=TRUE, 'mc.cores=1')
k <- writeLines(xx, "ee.csv")
close(fileConn)
```

## C.5 Pre-processing of Social Media Data

```
File4<-file("ll.csv")
yy <- readlines("ll.csv")
yy<-tm_map(yy,removeNumbers, lazy=TRUE, 'mc.cores=1')
yy<-tm_map(yy,stripWhitespace, lazy=TRUE, 'mc.cores=1')
yy<-tm_map(yy,removePunctuation, lazy=TRUE, 'mc.cores=1')
yy<-tm_map(yy,strip_retweets, lazy=TRUE, 'mc.cores=1')
yy<-tm_map(yy,removeWords,stopwords(english), lazy=TRUE, 'mc.cores=1')
fb <- writeLines(yy, "ll.csv")
close(file4)
```

## C.6 Data Cleaning

```
install.packages("tm")
install.packages("stopwords")
library(tm)
library(stopwords)

fn="f3.txt"
connnect <- file(fn,open="w")
tx1 <- readLines(connnect)
tx2 <- tolower(tx1)
tx3 <- removeWords(tx2,stopwords())
low <- writeLines(tx3, con = connnect, sep = "\n", useBytes = FALSE)
tx3
close(connnect)
```

## C.7 EE Computation

```
nrow(read.csv("agriculture-forestry.csv", header= T/F))
nrow(read.csv("mining-quarrying.csv", header= T/F))
nrow(read.csv("manufacturing.csv", header= T/F))
nrow(read.csv("electricity-gas.csv", header= T/F))
nrow(read.csv("water-sewerage.csv", header= T/F))
nrow(read.csv("construction.csv", header= T/F))
nrow(read.csv("wholesale-retail.csv", header= T/F))
nrow(read.csv("transportation-storage.csv", header= T/F))
nrow(read.csv("accommodation-food.csv", header= T/F))
nrow(read.csv("information-communication.csv", header= T/F))
nrow(read.csv("financial-insurance.csv", header= T/F))
nrow(read.csv("real-estate.csv", header= T/F))
nrow(read.csv("professional-scientific.csv", header= T/F))
nrow(read.csv("administrative-support.csv", header= T/F))
nrow(read.csv("public-administration.csv", header= T/F))
nrow(read.csv("education.csv", header= T/F))
nrow(read.csv("health-social_work.csv", header= T/F))
nrow(read.csv("entertainment.csv", header= T/F))
nrow(read.csv("household-activities.csv", header= T/F))
nrow(read.csv("extraterritorial-organization.csv", header= T/F))
nrow(read.csv("other-activities.csv", header= T/F))
```

## C.8 Dictionary of Financial Words Using Social Media

```
file1<-file("user.csv")
file2<-file("word.txt")

x<- readLines("user.csv")
y<- readLines("word.txt")
z<- grep(x, y, value = TRUE)
if(z == 1){
count <- count +1 }
k <- writeLines(count, "cc.csv")
close(file1)
close(file2)
```

## C.9 Detecting Non-Residents

```
File4<-file("user.csv")
File5<-file("country.txt")
x<- grep("location", readLines("country.txt"), value = TRUE)
if(x == 1){
count <- count +1 }
k <- writeLines(count, "non-resident.csv")
close(file4)
close(file5)
```

## C.10 Employee Computation

```
File5<-file("user.csv")
File6<-file("employee.csv")
x<- grep("student", readLines("user.csv"), value = TRUE)
y<- grep("house_wife", readLines("user.csv"), value = TRUE)
if(x == 1) break
else if (y==1) break
else {count <- count +1 }
k <- writeLines(count, "employee.csv")
close(file5)
close(file6)
```

# Appendix D

# Sample Dataset

## D.1 Web Media Dataset

```
Name,Address,Contact
Hotel Herocity,"3/A, Sadarghat Road ","031 612645, 619630"
Hotel Ovisar (Pvt) Ltd.,"Sea-Beach Road, Cox's Bazar ",0341-63061
Hotel Regent Park,"4/A, Jubilee Road, Chittagong ",031-2852072
Hotel Saint Martin Limited,"25, S. K. Mujib Road, Agrabad C/A ",725961-2
Hotel Abakash,Agrabad ,"8821548, 9899290"
Meridian Hotel & Restaurant,1367 CDA Avenue ,031 654000-1
Motel Shaikat,Chittagong ,880-31-209845
Mistimela,"1205, SK Mujib Road ",031-714832
Pagasai Restaurant,"Sajib Plaza, (1st Floor),12/A, Block -G,Halishahar Ho
Pizza Hut Bangladesh (Chittagong Outlet),"805/B, C.D.A Avenue, JB Complex
Restaurant (Ordinary Food),"Delight, Sadarghat ",616158
Swiss,"125, K.B. Fazil Kader Road. ",031 653096
Asian SR Hotel Chittagong Office,"291, Satation Road, Chittagong-4000, Ba
Duncan Products Ltd.,"Ispahani Building (1st Floor) Sheikh Mujib Road, Ag
Fried Peppers Restaurant,69 Agrabad C/A ,"723407, 2513030"
Genesis Hotel Raj Complex,"154, Kabi Nazrul Islam Road, Sadarghat ","031-
Hotel Agrabad,"P.O. Box # 147, Post Code # 4000, Agrabad ",031 713311-8
Hotel Al-Hera,"Hafiz Plaza 151-152, Kapasgola Road ","031 636947, 612364"
Hotel Atlantic Ltd,"Zilla Porishad Super Market, Court Road ",031 621460
Hotel Broadway,680 A. Jublee Road ,031 619795
Hotel Gateway,"82, Station Road ","031 634341, 0171861002"
Hotel Gateway,"82, Station Road ",031 634341
```

## D.2 Social Media Dataset

```
1 558523984664452,Masquer Rahman,0,Dhaka,0,CEAT AKKHAN LTD,
2 Elysa's Eid Shopping Done and she wear her one of Eid dress in the market
3 Good Morning from Bogra,,,,,,
4 Masquer Rahman checked in to Appollo Hospitals Dhaka with Torikul Udoy an
5 Roaming on Eid- with Masquer Rahman and 2 others,,,,,,
6 Team Lunch with Masquer Rahman and Mahmoodul at Alfresco Tejgaon,,,,,,
7 Good Luck Team Bangladesh,,,,,,
8 On the way to kishorgonj,,,,,,
9 Masquer Rahman eating ifter with Raisa Tasnim Rahman and 2 others at Adda
10 Alhamdullah. Thnx Raisa Tasnim Rahman for all hard work for preparing del
11 "Sharif Islam eating pizza with Masquer Rahman at Bella Italia, gulshan1"
12 CEAT Annual Picnic,,,,,,
13 Darjeling Mall,,,,,,
14 Birthday Celebration in office & Dinner with Family without my princess a
```

Generated using Postgraduate Thesis LaTeX Template, Version 0.97. Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

This thesis was generated on Saturday 8th February, 2020 at 3:48am.