Global Economic Prospects and Risk Factors

Parth Rustagi[1], Rohit Sroa[2], Priyanshu Sinha[3], Shikha Gupta[4]

[1,2,3]*Students, Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi*
[4]*Asst. Professor, Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi*

Email: rohitsroa@hotmail.com

Keywords:

Time Series Analysis, Deep Learning, ARIMA Model, LSTM Model, PROPHET Model, RMSE.

A B S T R A C T

Machine Learning and Data Science is the next big thing. This is fairly visible in the quote by Bill Gates that says "A breakthrough in machine learning is worth 10 Microsoft". It has the immense capability and potential to change the world. There is still a lot more to discover in this field and this is a paper attempt to at least cover some ground on those discoveries. The paper provides a comparative study of the different Time Series Analysis machine learning models using a real-world scenario. The scenario is to forecast the various risk factors associated with a country. A risk factor is defined as something that can harm the economic and political status of the country. We aimed to select some of the worst-hit countries of the world and predict the prospects of these countries on different parameters. The parameters chosen were Gross Domestic Product Per Capita Income (GDPI), Population, CO_2 damage and Life Expectancy. The study is a comparative study between ARIMA Model, LSTM Model, and FB Prophet Model.

Introduction

The 21st century has witnessed the biggest leaps in economy and technology. The advances in these fields have allowed various countries to develop at a tremendous rate with each country working to improve its economic, Political, and military status. A countries welfare is an amalgamation of the above-mentioned factors and many more. Each of these factors has a certain influence on the betterment of the country. If anyone of these factors overpowers the other it may lead the country into extinction. These factors must go hand in hand to ensure the proper development of the nation. It thus becomes essential to determine the associated risks of a country based on these factors. Our project aims to analyse various economic, political, and military factors to determine the associated risks of the country. The current system determines the risk factors based on the economic stability, by identifying rate of growth of per capita (GDP) and the FSI index of the country. The analysis of rate of growth per capita and the FSI index should provide us with the necessary information to assess the vulnerabilities of the countries. Our system further extends to utilise and analyse the degree of political severity to determine the political

stability and assess the extent of protests and violence. With the help of the political stability index we can determine the Likelihood of the politically motivated violence or terrorism that hampers the welfare of country.

Objectives

Objectives of the Project:

1.1 Predicting Economic stability by rate of growth of per capita GDP

1.2 Predicting risk factor of a country based on life expectancy at birth

1.3 Predicting factors based on adjust net saving: carbon dioxide damage in US$

1.4 Predicting risk factor of a country based on its subnational population

The objectives are discussed in detail as follows:

1.1 Predicting Economic stability by rate of growth of per capita GDP

Per capita GDP (Gross domestic product) is a metric that separates a nation's monetary yield for each individual and is determined by isolating the Gross domestic product of a nation by its populace.

There can be a couple of approaches to examine a nation's riches and flourishing. Per capita Gross domestic product is the most all-inclusive on the grounds that its segments are routinely followed on a worldwide scale, accommodating simplicity of count and utilization.

At its most essential understanding per capita Gross domestic product shows how much monetary creation worth can be ascribed to every individual resident. On the other hand, this means a proportion of public abundance since Gross domestic product market esteem per individual likewise promptly fills in as a thriving measure. (Zainab, Wani, & Bhat , 2018)

Per capita Gross domestic product is frequently examined close by Gross domestic product. Financial experts utilize this measurement for knowledge on both their own nation's home-grown efficiency just as profitability of different nations. Per capita Gross domestic product thinks about both a nation's Gross domestic product and its populace. In this manner, it tends to be essential to see how each factor adds to the general outcome and how each factor is influencing per capita Gross domestic product development.

1.2 Predicting risk factor of a country based on life expectancy at birth

By and large, another conceived can hope to live, if current demise rates don't change. In any case, the real age-explicit demise pace of a specific birth partner can't be known ahead of time. On the off chance that rates are falling, genuine life expectancies will be higher than future determined utilizing current demise rates. Future upon entering the world is one of the most habitually utilized wellbeing status markers. Additions in future upon entering the world can be credited to various components, including rising expectations for everyday comforts, improved way of life and better training, just as more noteworthy admittance to quality wellbeing administrations. This marker is introduced as an aggregate and per sex and is estimated in years. Future upon entering the world mirrors the general mortality level of a populace. It sums up the mortality design that wins over all age gatherings - kids and young people, grown-ups and the older. Future upon entering the world is gotten from life tables and depends on sex-and age-explicit passing rates. Future upon entering the world qualities from the Unified Countries relate to mid-year gauges, predictable with the comparing Joined Countries fruitfulness medium-variation quinquennial populace projections.

Methods used to appraise WHO life tables for Part States shift contingent upon the information accessible to survey youngster and grown-up mortality. Four nation classifications have been utilized for this update. In every one of the three cases, UN-IGME appraisals of neonatal, baby and under-5 death rates were utilized. 1) Nations with high HIV for which WPP2015 utilized Range to unequivocally display HIV mortality. The UN Populace Division has given unpublished evaluations of non-HIV mortality for these nations. 2) Nations with critical HIV plagues for which WHO has in the past unequivocally displayed HIV and non-HIV mortality patterns to plan life tables. These nations were not demonstrated utilizing Range for WPP2015. 3) Nations for which the WHO Mortality Information base held mortality information from indispensable enrolment (VR) frameworks for 75% or a greater amount of years since 1990. 4) Nations where interjected death rates from WPP quinquennial life tables were utilized straightforwardly to develop yearly life table's Overwhelming kind of measurements: Anticipated. (Ratakonda & Sasi, 2019)

1.3 Predicting factor of a country based on adjusted net saving: carbon dioxide damage in US$

Cost of harm because of carbon dioxide emanations from petroleum product use and the assembling of concrete, assessed to be US$30 per ton of CO2 (the unit harm in 2014 US dollars for CO2 transmitted in 2015) times the quantity of huge loads of CO2 radiated. World Bank staff gauges dependent on sources and strategies depicted in "The Changing Abundance of Countries 2018: Building a Manageable Future" (Lange et al 2018).

Carbon dioxide (CO2) is a scentless gas that is profoundly critical to life on Earth. CO2 is otherwise called an ozone depleting substance; an exorbitant focus can upset the normal guideline of temperature in the environment and lead to an unnatural weather change.

The grouping of CO2 has particularly expanded because of the modern insurgency and dramatic development in assembling exercises the world over. Deforestation, agribusiness, and petroleum derivative use are the essential wellsprings of CO2. As per the latest information from the Worldwide Carbon Venture, the best five nations that delivered the most CO2 are China, the U.S., India, Russia, and Japan.

CO2—otherwise called ozone depleting substances—has become a significant worry as environmental change turns into a greater issue. China is the world's biggest contributing nation to CO2 emanations—a pattern that has consistently ascended throughout the long term—presently delivering 10.06 billion metric huge loads of CO2. The greatest offender of CO2 discharges for these nations is power, remarkably, copying coal. Carbon dioxide discharges are the essential driver of worldwide environmental change. It's broadly perceived that to evade the most exceedingly awful effects of environmental change, the world requirements to direly lessen outflows. Be that as it may, how this duty is shared between districts, nations, and people has been an unending purpose of dispute in global conversations. This discussion emerges from the different manners by which emanations are looked at: as yearly outflows by nation; discharges per individual; recorded commitments; and whether they change for exchanged merchandise and ventures. These measurements can recount totally different stories

1.4 Predicting factor of a country based on its subnational population

Subnational Populace Information base presents assessed populace at the main authoritative level beneath the public level. A considerable lot of the information come from the nation's public measurable workplaces. Other information come from the NASA Financial Information and Applications Centre (SEDAC) oversaw by the Middle for Worldwide Geology Data Organization (CIESIN), Earth Foundation, and Columbia College. It is the World Bank Gathering's first subnational populace information base at a worldwide level and there are information constraints. Arrangement metadata incorporates strategy and the suspicions made. The Evaluation Agency computes subnational 5-year age/sex bunch populace appraisals and projections for the years 2000 through 2015, 2020, or 2025 for nations imparting two-sided endeavours to the U.S. Government as a component of PEPFAR. The PEPFAR program gives help to nations around the globe whose populaces experience the ill effects of a high pace of HIV disease. Our subnational populace information are predictable with the public projections from the U.S. Enumeration Agency's Global Information Base and are connected to advanced guides of the subnational regulatory zones. These items likewise are accessible through the DHS Program Spatial Information Store.

2. Methodologies

2.1 Time Series Analysis Forecasting

Definition of Time Series: An ordered sequence of values of a variable at equally spaced time intervals.

Time arrangement examination is a factual procedure that manages time arrangement information, or pattern investigation. Time arrangement information implies that information is in a progression of specific time-frames or spans.

The information is considered in three sorts:

- Time arrangement information: A bunch of perceptions on the qualities that a variable takes at various occasions.

- Cross-sectional information: Information of at least one factors, gathered at a similar point as expected.

- Pooled information: A blend of time arrangement information and cross-sectional information.

Time Series Analysis is used for many applications such as:

- Economic Forecasting
- Sales Forecasting
- Budgetary Analysis
- Stock Market Analysis
- Yield Projections
- Process and Quality Control
- Inventory Studies
- Workload Projections
- Utility Studies
- Census Analysis

This strategy predicts the one next period esteem dependent on the past and current worth. It includes averaging of information with the end goal that the nonsystematic segments of every individual case or perception offset one another. The outstanding smoothing strategy is utilized to anticipate the transient predication. Alpha, Gamma, Phi, and Delta are the boundaries that gauge the impact of the time arrangement information. Alpha is utilized when irregularity is absent in information. Gamma is utilized when an arrangement has a pattern in information. Delta is utilized when irregularity cycles are available in information. A model is applied by the example of the information. Bend fitting in time arrangement investigation: Bend fitting relapse is utilized

when information is in a non-direct relationship. The accompanying condition shows the non-direct conduct:

Subordinate variable, where case is the successive case number. Bend fitting can be performed by choosing "relapse" from the examination menu and afterward choosing "bend assessment" from the relapse choice. At that point select "needed bend direct," "power," "quadratic," "cubic," "backwards," "calculated," "dramatic," or "other." (Ratakonda & Sasi, 2019)

2.2 Arima Model

An ARIMA model is a class of measurable models for breaking down and estimating time arrangement information.

It unequivocally obliges a set-up of standard structures in time arrangement information, and as such gives a straightforward yet amazing strategy for making handy time arrangement conjectures.

ARIMA is an abbreviation that represents Auto Regression Integrated Moving. It is a speculation of the less difficult Auto Backward Moving Normal and adds the thought of combination.

AR: Auto Regression. A model that utilizes the reliant connection between a perception and some number of slacked perceptions.

•I: Integrated. The utilization of differencing of crude perceptions (for example deducting a perception from a perception at the past time step) to make the time arrangement fixed.

MA: Moving Average. A model that utilizes the reliance between a perception and a lingering blunder from a moving normal model applied to slack perceptions.

Every one of these parts are unequivocally indicated in the model as a boundary. A standard documentation is utilized of ARIMA (p, d, and q) where the boundaries are subbed with whole number qualities to rapidly demonstrate the particular ARIMA model being utilized.

The boundaries of the ARIMA model are characterized as follows:

P: The quantity of slack perceptions remembered for the model, likewise called the lag order.

D: The occasions that the crude perceptions are differenced, additionally called the degree of differencing.

Q: The size of the moving normal window, additionally called the order of moving average.

A straight relapse model is built including the predetermined number and sort of terms, and the information is set up by a level of differencing to make it fixed, for example to eliminate pattern and occasional structures that adversely influence the relapse model.

An estimation of 0 can be utilized for a boundary, which shows to not utilize that component of the model. Along these lines, the ARIMA model can be designed to play out the capacity of an ARMA model, and even a basic AR, I, or MA model.

Embracing an ARIMA model for a period arrangement accepts that the basic cycle that produced the perceptions is an ARIMA cycle. This may appear glaringly evident, yet assists with spurring the need to affirm the suppositions of the model in the crude perceptions and in the leftover blunders of figures from the model. (Salma , Hassan, & Sultana, 2020). The stats models library provides the capability to fit an ARIMA model. An ARIMA model can be created using the stats model's library as follows:

- Define the model by calling ARIMA () and passing in the p, d, and q parameters.

- The model is prepared on the training data by calling the fit () function.

- Predictions can be made by calling the predict () function and specifying the index of the time or times to be predicted.

The ARIMA model can be utilized to gauge future time steps. We can utilize the predict () work on the ARIMA Results object to make expectations. It acknowledges the record of the time steps to make expectations as contentions. These records are comparative with the beginning of the preparation dataset used to make expectations. On the off chance that we utilized 100 perceptions in the preparation dataset to fit the model, at that point the list of whenever venture for making an expectation would be indicated to the forecast work as start=101, end=101. This would restore a cluster with one component containing the expectation. We additionally would favor the gauge qualities to be in the first scale, on the off chance that we played out any differencing (d>0 while arranging the model). This can be determined by setting the typ contention to the worth 'levels': typ='levels. Then again, we can

dodge these determinations by utilizing the forecast t() work, which plays out a one-venture figure utilizing the model. We can part the preparation dataset into train and test sets, utilize the train set to fit the model, and create an expectation for every component on the test set. A moving gauge is required given the reliance on perceptions in earlier time ventures for differencing and the AR model. A rough method to play out this moving conjecture is to re-make the ARIMA model after each groundbreaking perception is gotten. We physically monitor all perceptions in a rundown considered history that is cultivated with the preparation information and to which novel perceptions are annexed every cycle.

2.3 LSTM (Long Short-Term Memory)

The Long Short-Term Memory network, or LSTM network, is a repetitive neural organization that is prepared utilizing Backpropagation through Time and beats the evaporating slope issue. Thusly, it very well may be utilized to make enormous intermittent organizations that thusly can be utilized to address troublesome succession issues in AI and accomplish cutting edge results. Rather than neurons, LSTM networks have memory obstructs that are associated through layers. A

square has parts that make it more astute than an old-style neuron and a memory for late arrangements. A square contains doors that deal with the square's state and yield. A square works upon an information arrangement and each entryway inside a square uses the sigmoid enactment units to control if they are set off, rolling out the improvement of state and expansion of data moving through the square restrictive.

There are three types of gates within a unit:

- Forget Gate: conditionally decides what information to throw away from the block.
- Input Gate: conditionally decides which values from the input to update the memory state.
- Output Gate: conditionally decides what to output based on input and the memory of the block.

Each unit is like a mini-state machine where the gates of the units have weights that are learned during the training procedure.
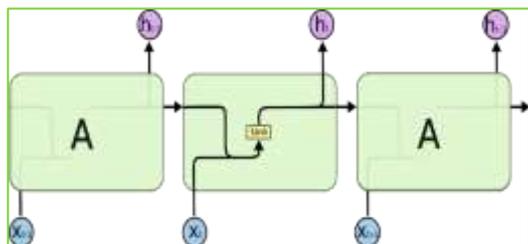
2.3.1   LSTM Gate Diagram

LSTMs additionally have this chain like structure, yet the rehashing module has an alternate structure. Rather than having a solitary neural organization layer, there are four, cooperating in an exceptionally unique way.

2.4 Prophet Model

The Prophet library is an open-source library intended for making figures for univariate time arrangement datasets. It is anything but difficult to utilize and intended to naturally locate a decent arrangement of hyper boundaries for the model with an end goal to make capable figures for information with patterns and occasional structure of course. Prophet is a method for gauging time arrangement information dependent on an added substance model where non-direct patterns are fit with yearly, week after week, and every day irregularity, in addition to

occasion impacts. It works best with time arrangement that have solid occasional impacts and a few periods of verifiable information. Prophet is hearty to missing information and movements in the pattern, and ordinarily handles exceptions well. Anticipating is frequently viewed as a characteristic movement from revealing. Detailing causes us answer, what was the deal? Gauging helps answer the following consistent inquiry, what will occur?

Prophet is especially useful for datasets that:

- Contain an extended time period (months or years) of detailed historical observations (hourly, daily, or weekly)
- Have multiple strong seasonality
- Include previously known important, but irregular, events
- Have missing data points or large outliers
- Have non-linear growth trends that are approaching a limit

Prophet is an added substance relapse model with a piecewise straight or calculated development bend pattern. It incorporates a yearly occasional part displayed utilizing Fourier arrangement and a week after week occasional segment demonstrated utilizing sham factors. (Bilal, et al., 2019)

Truly, great gauges have been trying to create. This brought about an extreme deficiency of experts who could convey figures with the degree of exactness needed to drive business choices. To reduce this gracefully hole and to make versatile anticipating drastically simpler, the Centre Information Science group at Facebook made Prophet, a gauging library for Python and R, which they publicly released in 2017. The expectation behind Prophet is to "make it simpler for specialists and non-specialists to make great figures that stay aware of interest." Prophet can deliver dependable and strong gauges (frequently performing in a way that is better than other regular estimating strategies) with next to no manual exertion, while taking into account the use of space information by means of effectively interpretable boundaries. In this formula, you'll figure out how to utilize Prophet (in Python) to tackle a typical issue: estimating an organization's day by day arranges for the following year. This lightweight model should fill in as an incredible manner to begin with Prophet, and will ideally start some motivation to plunge significantly more profound into the library's huge usefulness. This recipe is broken down into four main sections:

- Data Preparation & Exploration
- Box-Cox Transform
- Forecasting
- Inverse Box-Cox Transform

2.5 Stream lit Library

Stream lit is an open-source Python library that makes it simple to make and share excellent, custom web applications for AI and information science. In only a couple minutes you can manufacture and convey incredible information applications. A couple of the benefits of utilizing Streamlit apparatuses like Scramble and Carafe:

• It grasps Python scripting; No HTML information is required!

• Less code is expected to make a wonderful application

• No call-backs are required since gadgets are treated as factors

• Data storing improves and accelerates calculation pipelines.

For Data Mining and Predictive Analysis Process

•   Data   Understanding-Information understanding   depends   on   business understanding. Information is gathered at this phase of the cycle. The comprehension of what the business needs constantly will figure out what information is gathered, from what sources, and by what strategies. It consolidates the   phases   of   Information   Prerequisites, Information   Assortment,   and   Information Understanding from the Basic Strategy layout.

•   Modelling-Once   ready   for   use,   the information will be communicated through whatever   fitting   models,   give   important experiences, and ideally new information. This is the motivation behind information mining: to make information data that has importance and utility. The utilization of models uncovers examples and structures inside the information that give knowledge into the highlights of interest. Models are chosen on a part of the information   and   changes   are   made   if fundamental.

• Deployment In the arrangement step, the model is utilized on new information outside of the extent of the dataset. The new connections at this stage may uncover the new factors and

requirements for the dataset and model. These new difficulties could start modification of either business needs and activities, or the model and information, or both.

Procedure   for   Arrangement   on   Web Application

•   Model   Building-This   generally   includes parting the information into a preparation set and an approval set, attempting various mixes of calculations, and tuning their boundaries and hyper-boundaries.

• Deployment-The arrangement of AI models is the way toward making models accessible underway where web applications, undertaking programming   and   APIs   can   devour   the prepared model by giving new information focuses and producing forecasts.

•   Monitoring   and   Recognizability   Once underway, we can utilize the checking and discernibleness foundation to assemble new information that can be examined and used to make new preparing informational indexes, shutting   the   input   circle   of   nonstop improvement.

3. Result

To showcase the result, we have taken INDIA as the primary country among the 10. We have shown comparison between the 3 methods that we have described above, i.e. ARIMA Model, LSTM Model and Facebook's Prophet Model. Root Mean Square Error (RMSE) is the standard deviation of the residuals (expectation mistakes). Residuals are a proportion of how a long way from the relapse line information focuses are; RMSE is a proportion of how spread out these residuals are. At the end of the day, it discloses to you how thought the information is around the line of best fit.

Below is the comparison of RMSE graphs for each of the four factors for INDIA where the **first** bar represents **ARIMA Model**, **second** represents **LSTM** and **third** represents **PROPHET** Model.
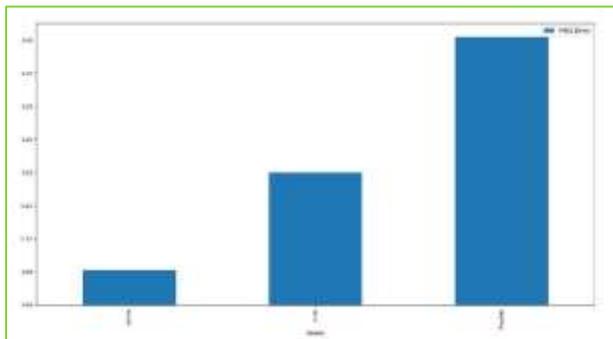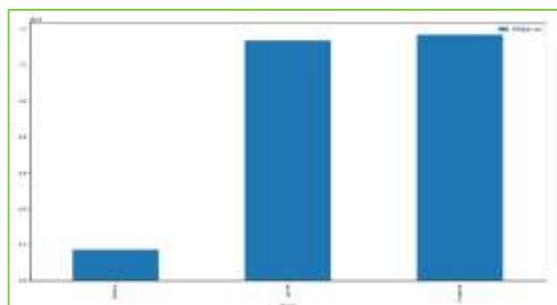

3.2. **RMSE Error of parameter: Life Expectancy Total**
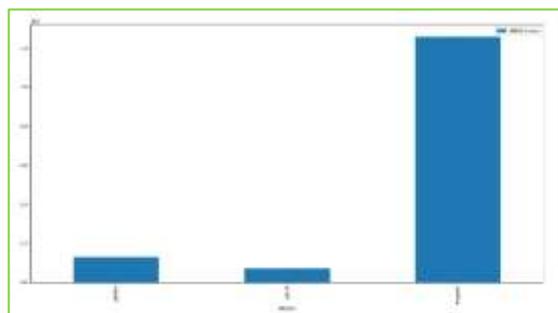
3.3. **RMSE Error of parameter: Subnational Population**

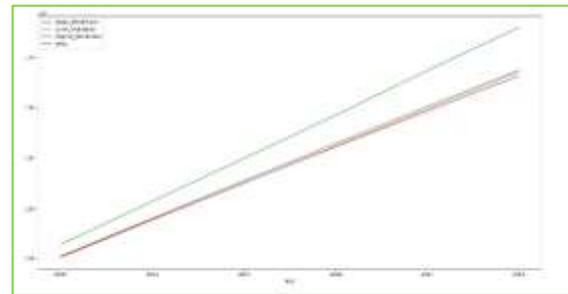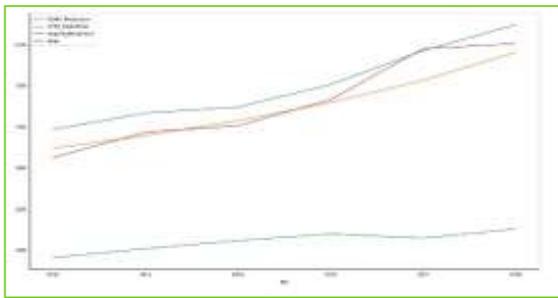**3.1 RMSE Error of parameter: GDP**
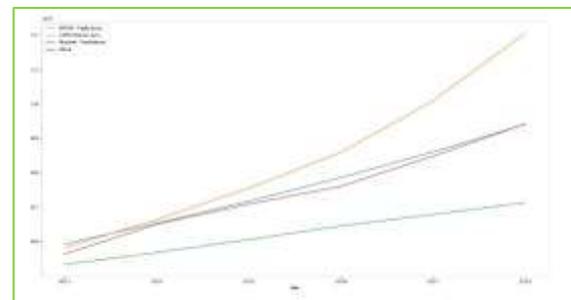

3.4 **RMSE Error of parameter: CO2 Damage**

Below are the comparison graphs. In the graph, the original value is represented by the **RED** line, for **ARIMA** Model **BLUE** line is used, for **LSTM**, **YELLOW** line is used and for **PROPHET** Model, **GREEN** line is used.
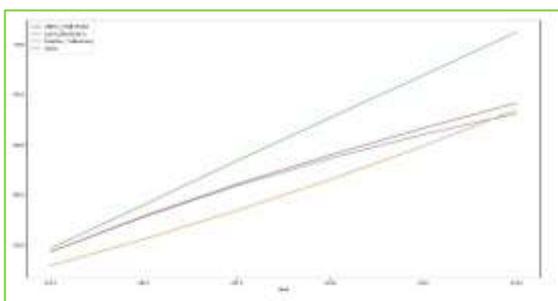

3.7   Comparison Graph of trained model: Subnational Population

3.5. Comparison Graph of trained model: GDP

3.8   Comparison Graph of trained model: CO2 Damage.

Conclusion

The paper provided a comparative Study of different Time Series Analysis models using real-life datasets. Different models were used, analysed and their performance was compared. This paper provides an insight as to how different machine learning models perform under real-life scenarios. It measures how

3.6   Comparison Graph of trained model: Life Expectancy.

accurate the results of the different models were. Data is the bread and butter of data science. No matter how good the model is it will always be restricted by the availability of data. In this scenario, any deviation in the result, the lack of data was a major contributor. Every country whether developed or not has factors that lead to its instability. The factors may not be visible now but they may come up anytime. Thus, every country should work to prevent these risks from erupting in the near future.

References

1) Bilal, M., Malik, N., Bashir, N., Majani, M., Hashin, T. A., & Gain, A. (2019). Profiling Social Media Campaigns and Political Influence: The Case of Pakistani Politics. *IEEE*.

2) Ratakonda, S. S., & Sasi, S. (2019). Seasonal Trend Analysis on Multi-Variate Time Series Data. *IEEE*.

3) Salma , U., Hassan, M. M., & Sultana, I. M. (2020). A Regression Model-Based Approach to Identifying Determining Factors for GDP Growth in Bangladesh. *IEEE*.

4) Shirota, Y., Presekai, A., & Sari, F. R. (2019). Visualization of Time Series Change on GDP Per Electricity by Provinces in Indonesia. *IEEE*.

5) Zainab, T., Wani, A. Z., & Bhat , A. M. (2018). Scientific Research in Relation To Gross Domestic Product (GDP) A Comparative Study of China and India. *IEEE*.