Analysis of Gender Bias in the Employment Sector

Srihaasa Pidikiti
srpi9270@coloardo.edu
CSCI 5502 - srpi9270

Amatullah Sethjiwala
amse5199@coloardo.edu
CSCI 5502 - amse5199

Xuefei Sun
xusu2102@coloardo.edu
CSCI 5502 - xusu2102

Preethi Vijai
prvi9376@coloardo.edu
CSCI 5502 - prvi9376

ABSTRACT

Gender bias is greatly widespread in the education and employment sectors. In both sectors women are less represented and highly discriminated against in comparison to men. In the employment sector, women tend to receive less pay and benefits in comparison to their male colleagues. The main goal of our research is to do an in-depth analysis of gender discrimination in the employment sector. First, we define a socio-economic index, called the Gender Employment Parity Index (GEPI) that identifies gender bias in the employment sector for 187 countries. We analyse this index based on the different geographical locations and varying income groups to identify any interesting patterns or observations. Furthermore, we examine the correlation between various employment and education factors to determine the effect bias in education has on employment. Lastly, we predict future trends of employments to aid government in formulating better policies.

KEYWORDS

Data mining, gender bias, correlation, data analysis

ACM Reference Format:
Srihaasa Pidikiti, Amatullah Sethjiwala, Xuefei Sun, and Preethi Vijai. 2020. Analysis of Gender Bias in the Employment Sector. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Gender bias is prevalent in each and every sector of employment. Women have lower wages, have fewer chances of promotions and are given lower ranked positions with respect to their opposite sex. Additional factors such as: lack of appropriate workplace facilities, unsafe travel, poor family-work life balance and fear of sexual harassment aid in further strengthening these disparities present between the male and female population in the employment sector. Looking at the education sector similar challenges are faced by women, especially by women pursuing higher education. Interestingly, statistics reveal better performance of the enrolled female population in comparison to enrolled male population in all levels of education. A few factors that contribute to this prejudice are- insufficient funding for women from low income and minority families, differences in their secondary education curriculum and a push towards more "feminine" courses like arts and humanities.
Gender inequality incurs a loss of up to USD 12 trillion or 16% of global income [6]. An immediate effort should be taken to mitigate the bias and to strive for equality irrespective of gender. This requires a proper plan by the government aiming to curb discrimination at the root level. However, due to the widespread diversity between men and women based on their social and economic background, it is a difficult challenge to formulate policies that provide women equality in all fields.
Our work aims to mitigate gender biases seen in the workforce by analyzing in detail the various factors responsible for such discriminations. The work has been split into the following parts :

- To aid in creating fair and smart policies, we define a Gender Employment Parity Index (GEPI) that identifies gender bias in three employment sectors - Agriculture, Industry and Services. This index is similar to the Gender Parity Index (GPI) which is a socioeconomic index defined by UNESCO to identify gender bias in access to education. We study this index to find key factors and insights depending on the multiple geographical areas and different income levels.
- We explore and investigate the various factors that contribute to these gender disparities in each country.
- To evaluate the impact gender inequality in education has on employment, we investigate the correlation between different employment and education factors.
- We forecast future employment trends in the agriculture, industry and services sector for the next five years based on gender. Using these predicted values, we also calculate the future GEPI for all the countries. We hope that these predictions will help countries prepare for what's coming next and also help them formulate better strategies to diminish gender disparities in the employment sector.

2 RELATED WORK

We start by reviewing prior work in the field of data analysis to study gender bias in various fields.
In the context of STEM Chopra et al. [5] study gender differences using various data mining approaches. They take on a data-intensive approach to research gender disparities in Science, Technology , Engineering and Mathematics (STEM). It is well noted that women are underrepresented in STEM degrees: only 23% of women with

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/1122445.1122456

high mathematics scores seek STEM degrees, compared to 45% of men with the same scores. [7] Combining deep learning, data mining and statistical methods they measure the gender gap for women in STEM. Since real world data is used in addition to state-of-the-art data analysis techniques, in contrast to previous works, they provide us with new perspectives. The papers findings suggest that women have different motivations for applying to an engineering program than men. Using neural nets and text mining approaches they found out that when asked the question "Why are you interested in engineering?", men and women have varying responses. Men tend to use words such as 'Extracurriculars' and 'Childhood Dreams' significantly more than women. In their decision to pursue a career in STEM, women mention personal and familial factors. Women demonstrate a greater desire to contribute to society in general. The paper also analyses gender differences in cooperative work places. Their results indicate that 86% of IT jobs are held by men and 68% of health jobs are held by women. Finally their work suggests that women are less likely to be involved in entrepreneurial practices as compared to men.

Petrovska and Goldberg [13] did a study on implicit biases and gender identity in the Department of Computer Science at the Technical University of Munich. The goal of this research was to examine whether and how often the female students in the department are influenced in their academic life by possible gender bias. The authors created a questionnaire for the department with various types of questions. Some questions were directed at capturing personal opinions and how much the individual fits with his or her own gender, while others were asked to understand the student's understanding of gender equality problems in the departments. Their findings suggest that female students and experts in the field of computer science will change their conduct in order to be sufficiently acknowledged and excel in a male-dominated environment. The findings also indicate that each individual more readily identifies computer science with their own gender, but males do so more than females do.

Cassidy R. Sugimoto [19] and her colleagues present a bibliometric analysis confirming that gender imbalances persist in research output worldwide. The analysis shows that, Globally, women account for fewer than 30% of authorships, while men account for slightly more than 70% authorships. The analysis also shows that female collaborations are more domestically oriented compared to the men. On the top of that, they discovered that the difference of scientific impact between women and men. As long as a woman was one of the authors, the paper attracted fewer citation. Their data suggest programmes fostering international collaboration for female researchers might help making a country more scientifically competitive. Unfortunately, it can be very hard to achieve due to the local and historical force on gender disparities.

Different from previous works, Mohsen Jadidi[10] and his team analyzed the temporal evolution of collaborations in one entire computer science field and compared the structural position of men and women. They explored the gender-specific differences in collaboration patterns of more than one million computer scientists over 47 year. They investigated how these patterns vary over time and the age that when they achieve success in their career. To analyse the evolution of gender disparities in the computer science community between 1970 and 2015, they compared the number of male and female scientists that stopped publishing (dropouts), number of publications per author (productivity), collaboration patterns and scientific success of male and female scientists. They found that the dropout rate of women is higher than men. Women are less productive on average because they have a smaller fraction of senior authors.

The paper proposed by Adele H. Marshall, Mariangela Zenga and Aglaia Kalamatianou [2] analyses gender differences according to two new indicators in higher education. One indicator represents the length of studies beyond the minimum requirement and the other represents a harmonised graduation mark. The data mining techniques uses survival trees and multivariate regression trees which lets us identify the importance of such influencing variables to be measured and illustrates these in a simple to view tree structure. The Multivariate Regression Tree approach allows us to consider more than one continuous outcome variable so we were able to consider the two proposed indicators simultaneously. The insights from this analysis are that gender has an important role and women outperform with respect to these new indicators by taking less length of studies, less graduation time with higher performance, controlling also for other student characteristics. The modified gender parity index enriched the results in all stages. In order to conduct the investigation, this paper reports on the analysis of two individual level administrative data sets derived from social sciences-oriented departments of an Italian and a Greek university. A common feature of these data sets, is that there is a minimum time for graduation (threshold for graduation), but there is no maximum for graduation thus provoking the use of censoring to analyse the data, and the statistical methodology. The purpose of this paper is to investigate differences between male and female students progress through an analysis of length of studies and to measure gender difference in student performance graduation mark.

The analysis by Adva Gadoth and Jody Heymann [3] utilizes cross-sectional indicators from United Nations, and they have constructed global indices of gender parity in education and work, for international comparison. Multivariable regression has been performed to assess relationships between gender parity index scores and national mortality rates or life expectancy indicators. This study extends past research by examining actionable areas of gender equality and their impact on both male and female survival. While longitudinal research is needed to examine both causality and mechanisms. Their findings suggest longevity gains for both women and men, and for all children through reduced maternal mortality, where greater parity in school and work is exhibited.

3 DATA COLLECTION AND PREPROCESSING

The data has been collected from the data sets available at World bank. Specifically we collected gender based employment data with respect to factors like- percentage of contributing family workers, percentage of employment in Agriculture, Industry and Services sector, labor force participation rate and percentage of population employed. For all these above mentioned employment factors data is considered for 187 countries from 1991 to 2020. To better understand the impact of demographics or economy of a country on its employment factors, we also gathered data about the income group and the region a country belongs to. We also assembled gender

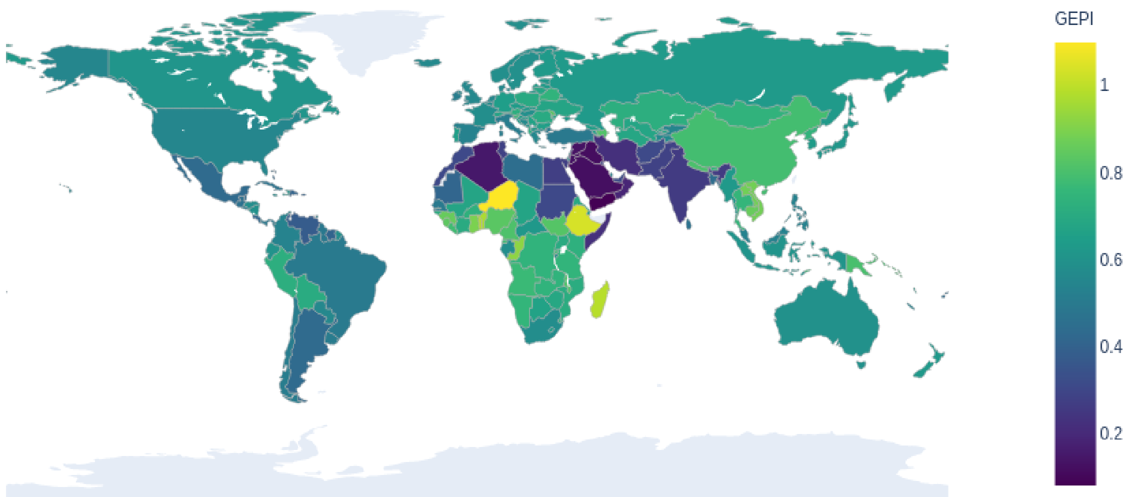

Figure 1: Average GEPI Heatmap from the years 2016 to 2020

based educational statistics from the data sets at World Bank based on factors such as- percentage of children out of primary school, literacy rate in the adult population, literacy rate in youth population of ages 15 to 24, percentage of primary completion rate and progression rate in school. For all these above mentioned employment factors data is considered for 187 countries from 1991 to 2020. All the data above was cleaned. Missing values were removed and prepossessed.

4 GENDER EMPLOYMENT PARITY INDEX

4.1 Methodology

The Gender Employment Parity Index (GEPI) was calculated as an average of three factors :

- **Agriculture Employment Parity ratio:** The ratio of the percentage of total population that is female and employed in agriculture to the percentage of total population that is male and employed in agriculture sector. The agriculture sector consists of activities in agriculture, hunting, forestry and fishing.[1]
- **Industry Employment Parity ratio:** The ratio of the percentage of total population that is female and employed in industry to the percentage of total population that is male and employed in industry sector. The industry sector consists of mining and quarrying, manufacturing, construction, and public utilities.[1]
- **Services Employment Parity ratio:** The ratio of the percentage of total population that is female and employed in services to the percentage of total population that is male and employed in services sector. The services sector consists of wholesale and retail trade and restaurants and hotels; transport, storage, and communications; financing, insurance, real estate, and business services; and community, social, and personal services. [1]

We choose these three factors owing to the fact that any country's economy is commonly categorised into three groups- primary, secondary, tertiary and definitions of agriculture, industry and services sectors align with these three groups respectively. We assign equal weight to each of the three factors since we wanted to accommodate the variations in the countries with respect to the three major income-generating sectors. Our hope is that this index will aid countries in understanding the level of gender bias present within the employment sectors and will also push countries in doing better for their female population.

4.2 Evaluation

4.2.1 ***Analysis of GEPI for Different Countries :*** The Heat map shown in Figure 1, is plotted with the average GEPI values over 5 years from 2016 to 2020 for each country. It is evident from the heat map that a majority of the countries in the world have a lesser female population employed compared to male population. In total, 8 countries have average GEPI value between 0 and 0.2. 17 countries have average GEPI value between 0.2 and 0.4. 68 countries have average GEPI value between 0.4 and 0.6. 73 countries have average GEPI value between 0.6 and 0.8. Lastly, 21 countries have average GEPI value between 0.8 and 1. We further conducted analysis of GEPI for different regions and income groups of countries to identify any common patterns with respect to GEPI over the years.

4.2.2 ***Analysis of GEPI for Different Regions :*** World Bank classifies a country into one of the seven regions based on the geographical location of the country. From Figure 2 it is clearly seen that no country has equal proportion of men and women employed. Middle East and North African countries have the smallest index indicating the greatest amount of gender bias present. Countries in North and South America and Europe have higher indices which goes with the norm.
The region with the highest index is the Sub Saharan Africa. According to a research, the top five countries with the maximum

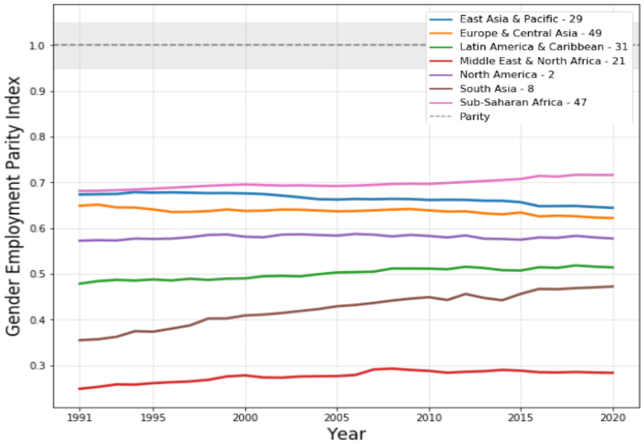

Figure 2: GEPI over the years for the seven geographical regions

number of women in the workforce are located in Sub-Saharan Africa. While this signifies progress, many women in this region are employed informally making them more likely to be underpaid, exploited and impoverished as well as having minimal job security. Of the women employed in Sub-Saharan Africa, 74 percent of them work in the informal sector which is significantly greater than the 61 percent of men in the region who are informally employed [12].

4.2.3 ***Analysis of GEPI for Different Income Groups :*** World Bank classifies each of the country into four types : High, Low, Upper middle and Lower middle based on its GNI percapita income in 2019. High income countries have percapita income of $12,536 or more. Low income countries have percapita income of $1035 or less. Lower middle income countries have percapita income between $1036 and $4045 whereas upper middle countries have it between $4046 and $12535. Out of the 187 countries whose GEPI is computed, 60 countries belong to High income group, 29 countries belong to Low income group, 48 countries belong to Lower middle income group and 50 countries belong to Upper middle income group.

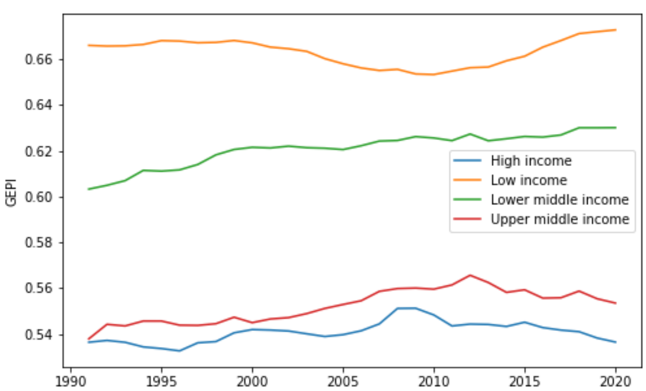

Figure 3: GEPI over the years for the four income group countries

Based on our observations of changes in GEPI over the years with respect to each of the income groups as shown in Fig. 3, we identified that low income countries have considerably less gender disparities in employment compared to the other three income groups. In contrast high income and upper middle income countries performed poorly with respect to GEPI over the years.

5 ANALYSIS OF GENDER BIAS IN THE VARIOUS FACTORS OF EMPLOYMENT

5.1 Methodology

We examine gender disparities in the following factors :

- Employment to Population Ratio,
- Employment in the Agriculture, Industry and Service sectors based on geographical location and income groups
- Contributing family workers.

5.2 Evaluation

5.2.1 ***Gender Bias in Employment :*** We use the Employment to Population Parity ratio for this analysis. The ratio is defined as the proportion of a country's population that is female and employed to the proportion of the country's population that is male and employed. Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period (i.e. who worked in a job for at least one hour) or not at work due to temporary absence from a job, or to working-time arrangements. Ages 15 and older are generally considered the working-age population [1].

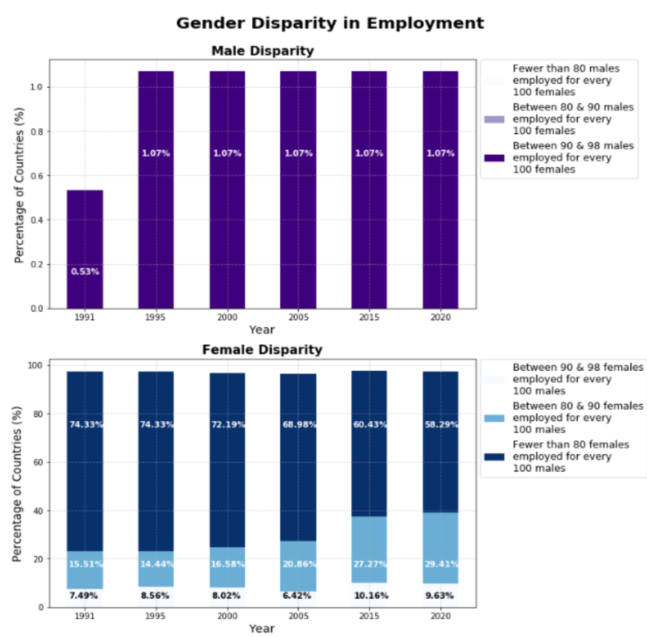

Figure 4: Male and Female Disparity in Employment

As seen from Figure 4 very few countries have achieved parity in employment based on gender. Majority of the countries (60%) have more men employed as compared to females, but on the positive side the percentage has been decreasing over the years. A possible reason for such disparity would be highly skewed Industry Employment Parity Ratio. Women tend to prefer jobs in the service sector as compared to the industry sector. Overall, the plots indicate that more incentives and laws need to be passed in all countries to make it easier for women to get a job.

5.2.2 ***Gender Bias in Different Sectors of Employment :*** As seen from the Figure 5 the industry sector has the highest gender bias suggesting that there aren't enough incentives provided to women to participate in such jobs. Additionally, women are much more likely than men to work in professional and related occupations as compared to labor intensive occupations [11].

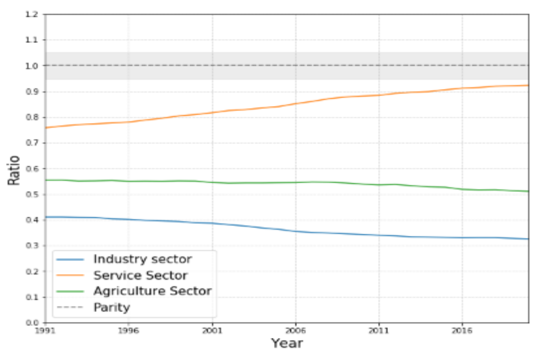

Figure 5: Parity Ratio of Different Sectors of Employment

The services sector has displayed an increasing trend towards gender parity indicating that more and more women are working and that the government is passing laws to help with this change. The differences in occupations in which women and men work are just one factor indicating that much more progress needs to be made before women can achieve equality in the workforce.

5.2.3 ***Gender Bias in Different Sectors of Employment based on the Regions :*** From Figure 6 we can observe a decreasing trend in the industry sector for all regions implying less and less women are inclined to work in this sector. To promote gender equality, the industry sector for all countries need to start making changes in their policies and create a female-friendly working environment. For the agriculture sector, South Asian and Sub-Saharan African countries have relatively less gender disparities. Although these regions have agriculture as their primary source of income, their parity ratio's show that more men are employed as compared to women. North American and European Countries dominate the services sector and in recent years have more females employed than male. Middle East and North African countries have low ratios irrespective of the sectors indicating there exist large gender bias in these regions when it comes to employment.

5.2.4 ***Gender Bias in Different Sectors of Employment based on the Income Groups.*** We analysed the temporal change of each of the three employment parity ratios that contribute towards GEPI calculation- Industry employment parity ratio, Agriculture employment parity ratio and Services employment parity ratio as shown in Fig. 7. Irrespective of the income groups, the Services employment parity ratio tends to be increasing over the years.Among all

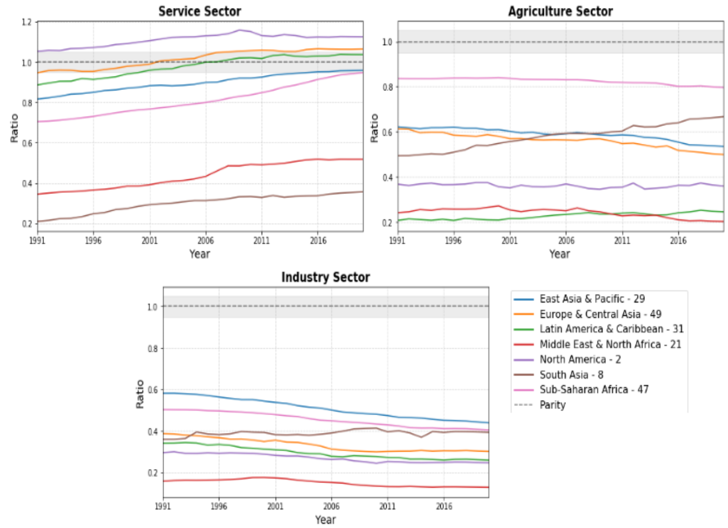

Figure 6: Services, agriculture and industry employment ratios over the years for the seven geographical regions

the income groups, high income group always employed a greater percentage of female population in services since 1990, followed by Upper Middle Income, Lower Middle and low income group countries in the same order.

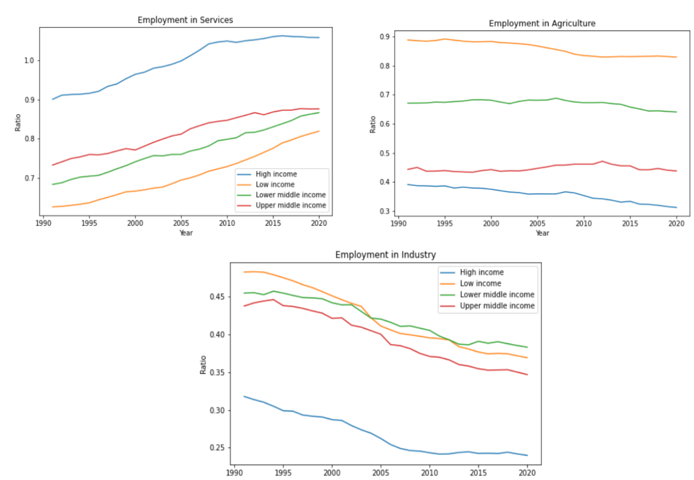

Figure 7: Services, agriculture and industry employment ratios over the years for the four income group countries

There is no specific increasing or decreasing trend over the years for all the four income groups, with respect to the Agriculture employment parity ratio. Among the four income groups, low income group always employed a substantial percentage of female population in agriculture since 1990, followed by Lower Middle Income, Upper Middle and high income group countries. Agriculture is still manual labor dependent in low income countries whereas in high income countries it is heading towards a complete machine dependent sector. Moreover, in low income countries, there is a

trend of male population in rural areas moving to urban areas for employment in industry and services sectors causing the increased active participation of women in agriculture sector [4]. These could be the possible reasons for a higher Agriculture employment parity ratio in low income countries
Regardless of the income groups, there is a decreasing trend over the years for the Industry employment parity ratio and this ratio is considerably lower in comparison to Services and Agriculture ratios since 1990. Among the four income groups, high income group always employed a lower percentage of female population in agriculture since 1990, whereas Lower Middle Income, Upper Middle and low income group countries have a comparatively higher ratio.

5.2.5 ***Gender Bias in Contributing Family Workers***. To investigate more into the employment sectors, we analyzed the contributing family worker data from 187 countries. We calculate the average contributing family worker employment rate for female and male between 1991 and 2020. On the Figure 8 and Figure 9, both contributing family worker employment rate are decreasing and it is exactly aligned with our previous result that the overall employment rate over the world is decreasing. By comparing these two figure together,we can tell that male usually have a higher contributing family worker rate which says it is more common for male to be the main family worker over these 33 year.

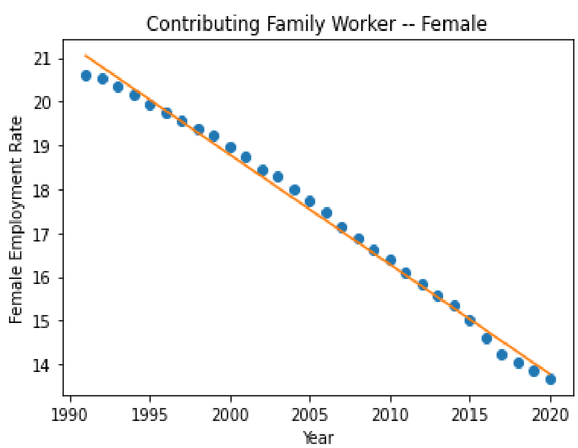

Figure 8: Contributing Family Worker Female

We also analyzed the the contributing family worker ratio over the years which is the calculation of average male family worker rate divided by the average female family worker rate for each year. On the Figure 10, the contributing family worker ratio is increasing over the years indicating that it is going to be more and more common for male to be the main family worker rather than the female. As the trend we observed here, it is easy for us to predict that the gender bias in contributing family worker will continue to increase in the future. It might require the government or some world agency to adopt some methods helping the women to get employed.
On addition to the analysis above, we further analyzed contributing family worker ratio over the countries. If the contributing family

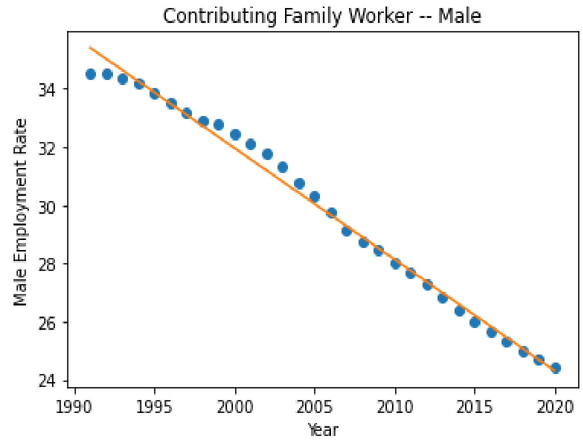

Figure 9: Contributing Family Worker Male

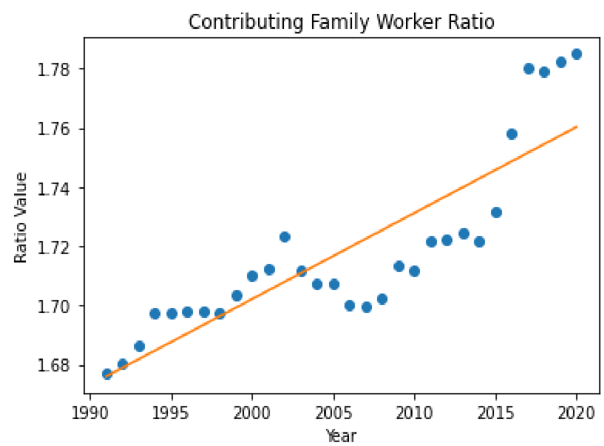

Figure 10: Contributing Family Worker Ratio Over Years

worker ratio is less than 1, more female function as the main family worker. And if it is greater than 1, it means the male is more common to be the family worker. On the Figure 11, Only less than 14% countries has a contributing family worker ratio less than 1 and over than half of these counties has a contributing family worker ratio over than 2. Considering our database includes a lot of developing countries, the result looks relatively reliable and trustworthy. Usually, the gender bias in the employment is more tend to be obvious in developing countries rather than the developed country.

6 CORRELATION ANALYSIS BETWEEN EDUCATION AND EMPLOYMENT

6.1 Methodology

The following correlation are computed between the education and employment.

- Correlation between Literacy Rate and Employment Rate Based on the Region and Gender
- Correlation between GPI and GEPI.

Figure 11: Contributing Family Worker Ratio Over Countries

6.2 Evaluation

6.2.1 ***Correlation between Literacy Rate and Employment Rate Based on the Region and Gender :*** We analysed the correlation between literacy rate and employment rate based on the geographical locations since the countries from the same region might share a lot of similarities on the background, culture and economy.

The countries from different regions might have other factor greatly affecting the correlation between employment rate and literacy rate. For example, the developed countries might have a very low agriculture employment rate because of the use of automatic seeder. While the developing countries might have a much higher agriculture employment rate with a similar literacy rate. We are expecting to see a similar trend on the employment rate by the change of literacy rate on the countries from the same region.

From Figure 14, we can tell that there is a positive correlation between literacy rate and the services employment rate. For most of regions, the services employment rate tends to increase when the literacy rate increases. The average correlation coefficient is around 0.6. We observed that the pattern is more obvious when the literacy rate is lower than 80%.

Europe and Central Asia is one exception here in the services sector. There is a very weak correlation between literacy rate and services employment rate with only 0.235 correlation coefficient for female and 0.215 for male. But we also noticed that countries in this region usually have a very high literacy rate which is over than 90%. So, we assume that the employment rate is more dependent on the country type after literacy rate achieves some threshold. For example, if people in this world all have a high school degree, the employment rate for one specific industry would more dependent on some other factors instead of the literacy rate.

We also noticed a negative correlation between employment rate and literacy rate for agriculture sector with an average correlation coefficient value of 0.62.

For industry sector, no obvious pattern or correlation was found. On addition to all the findings above, we also noticed with the same literacy level, the female employment rate is usually higher than the male employment rate in the services sector. On the Figure 14, we plotted best fit line for all the regions and both genders. The

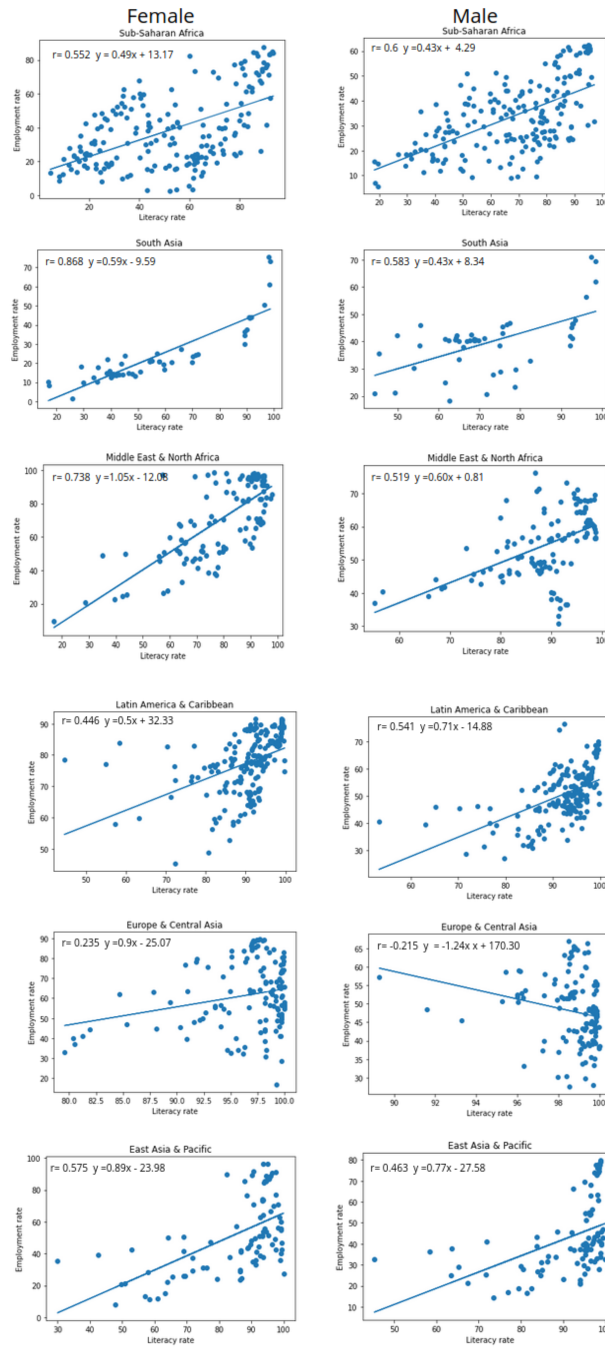

Figure 12: Correlation between Services Employment and Literacy Rate

function of best fit line is also annotated on each graph. The slope of the best fit line is higher for female than male which means with the same literacy level, the female workers prefer work in the services sector than the male. Similar pattern was found in the agriculture, the male workers prefer to work in the agriculture sector than the female with the same literacy rate.

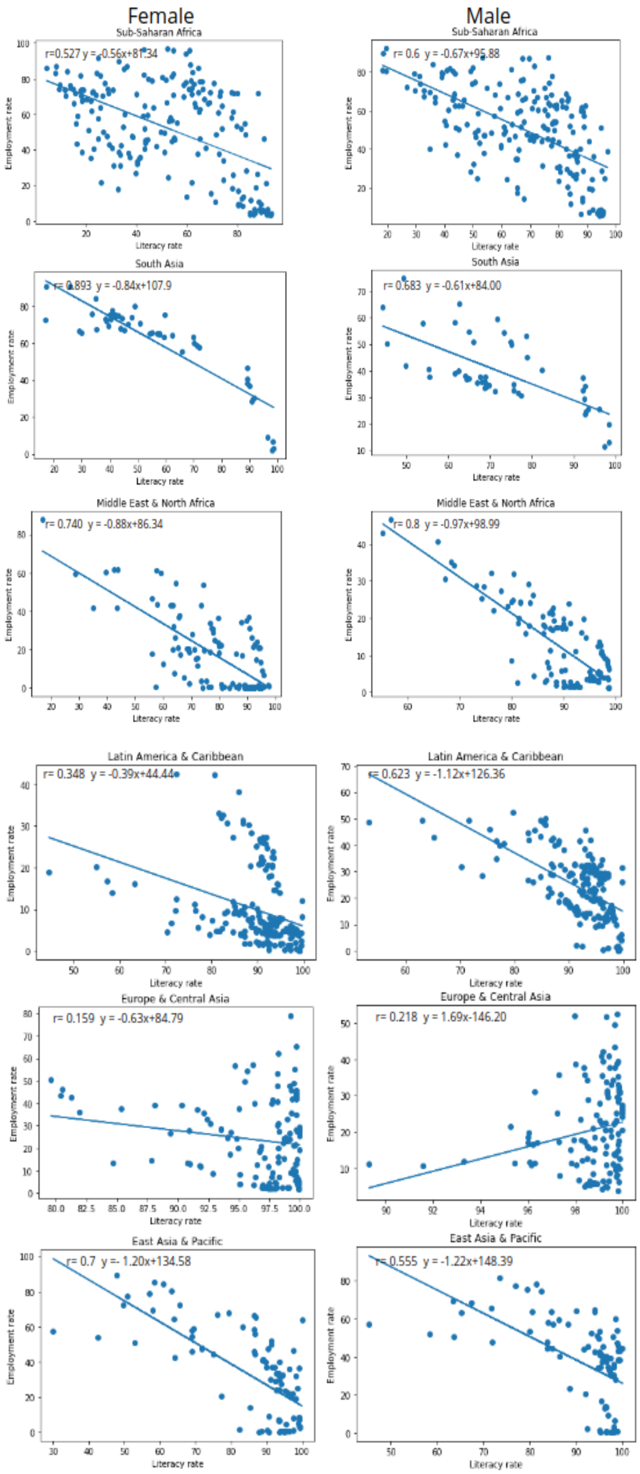

Figure 13: Correlation between Agriculture Employment and Literacy Rate

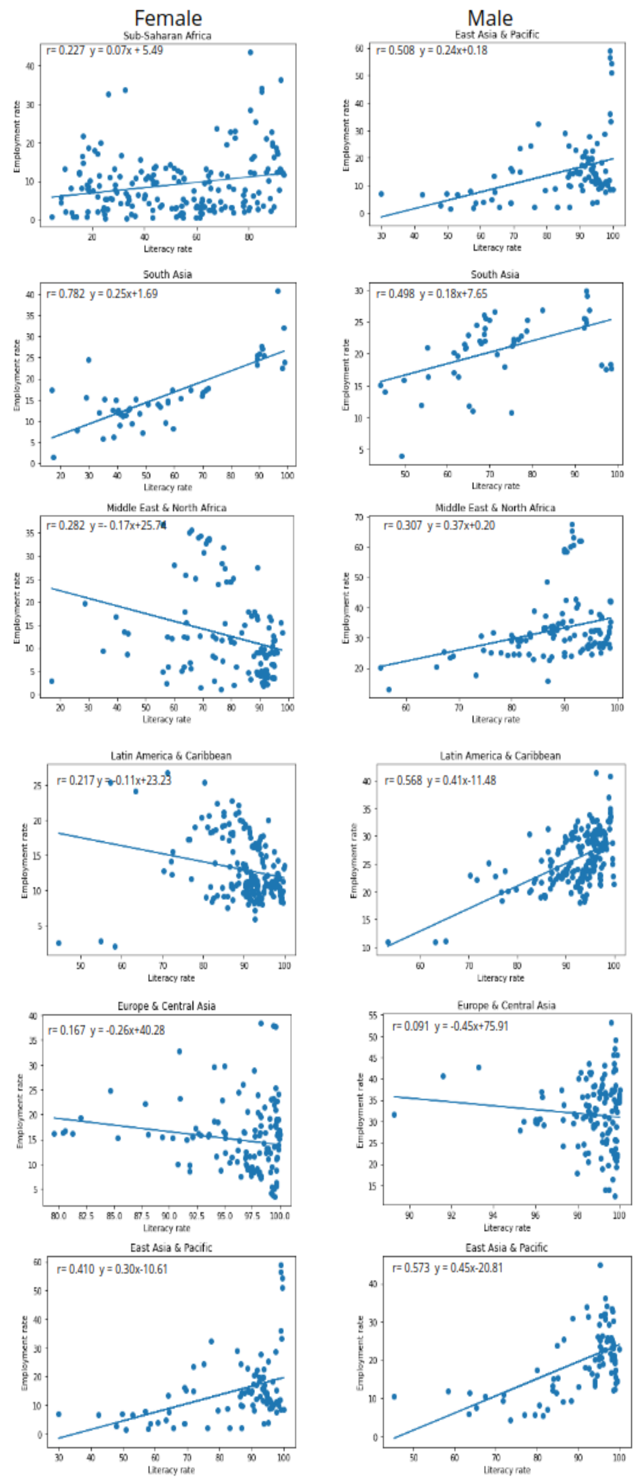

Figure 14: Correlation between Industry Employment and Literacy Rate

6.2.2 *Correlation between GPI and GEPI.*

On an average, a comparison across the regions(shown in the plots above) shows a positive correlation between GPI and Gender Employment Parity Indes. The correlation matrix is used to verify the features which are strongly correlated. From the above plot it is evident that GPI and GEPI are strongly correlated with a correlation coefficient of 0.85 and literacy vs employment is correlated with a correlation coefficient of 0.67. The gender parity index is positively correlated to the gender employment. The correlation is higher in the East Asia & Pacific region and Middle East & North Africa, where as lowest in South Asia. This could be an indicator of how the number of women in workforce alters the GPI in regions. It points to the fact that in some regions, women tend to be a part of lower wage public sectors.

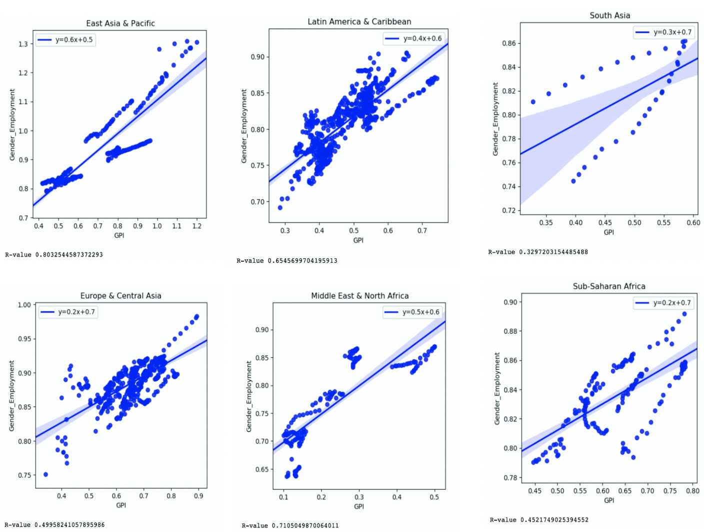

Figure 15: Correlation between GPI and GEPI

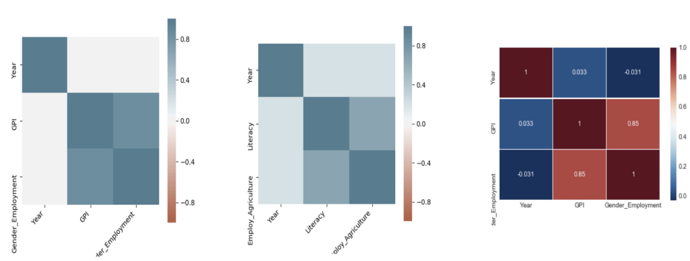

Figure 16: Correlation matrix

7 TIME SERIES PREDICTION MODELS FOR DIFFERENT SECTORS OF EMPLOYMENT

7.1 Methodology

The following models were utilized to predict the gender employment ratios in the Agriculture, Industry and Service sectors for the next 5 years (2021-2025):

- Simple Exponential Smoothing (SES) model
- Double Exponential Smoothing (DES) model
- AutoRegressive Integrated Moving Average (ARIMA) Model
- Holt's Linear Trend Model

Using these estimated employment ratios, we calculated the corresponding GEPI for each of the 187 countries. These forecast values can aid government officials in understanding where their countries are headed towards in terms of reducing gender bias and formulate action plans accordingly.

For each country, the training and testing data are split in 83:17 ratio i.e. 25 years of data (1991-2015) is reserved for training the model and 5 years of data (2016 - 2020) is utilized for testing purposes. The following metrics were chosen for evaluating the model's performance in each employment sector:

- Root mean square error (RMSE)
- Mean average percentage error (MAPE)
- Mean average error (MAE)

These metrics are calculated as the total average of error rates for each of the 187 countries in consideration.

7.1.1 *Simple Exponential Smoothing (SES) Model :*

In Simple Exponential Smoothing forecast, data is considered to not follow any trends in particular or exhibit any kind of seasonal effects [9]. Weighted average approach is followed in forecasting the values. Higher weights are assigned to the most recent values in comparison to the past values for this purpose. Since, some of the countries do not follow any particular trend and also some countries do follow trend only over a certain period of time and not the entire time range considered for the study, we decided to use this model in general to forecast the future employment ratios.

7.1.2 *Double Exponential Smoothing (DES) Model :*

As discussed previously, we do observe certain group of countries following either increasing or decreasing trends with respect to the employment ratios.In order to account for these data trends, Double Exponential Smoothing is a better choice over Simple Exponential Smoothing technique. Hence, this model has been considered for the study to understand the effect of trends on forecasts. In this approach, initially a weighted average approach is followed to predict the trend parameter for the year in consideration. This trend parameter is further used to estimate the value similar to Simple Exponential Smoothing.[8]

7.1.3 *AutoRegressive Integrated Moving Average (ARIMA) Model :*

ARIMA is a forecasting algorithm based on the assumption that past data alone can be used to estimate future values of the time series. In particular, it is a class of models that 'explains' a given time series on the basis of their own past values, that is, their own lags and lagged prediction errors, such that future values can be calculated using the equation [14]. With ARIMA models, any

Table 1: Evaluation Metrics for the Four Models

	Agriculture Sector			Industry Sector			Service Sector		
	RMSE	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*
SES Model	0.02716	0.07269	0.02553	0.01631	0.05122	0.01504	0.02786	0.03254	0.02592
DES Model	0.03353	0.08885	0.03063	0.02353	0.06839	0.02173	0.02904	0.03308	0.02622
ARIMA Model	0.03086	0.07921	0.02872	0.02086	0.06835	0.01903	0.02984	0.03755	0.02694
Holt's Model	0.03423	0.09440	0.03145	0.02235	0.06429	0.02058	0.02898	0.03313	0.02628

Table 2: Average of the Evaluation Metrics across all three factors

	RMSE	MAPE	MAE
SES Model	0.02378	0.05215	0.02216
DES Model	0.0287	0.06344	0.02619
ARIMA Model	0.02719	0.06170	0.02490
Holt's Model	0.02852	0.06394	0.02610

'non-seasonal' time series that displays patterns and is not a random white noise can be formulated. Since our data is non-seasonal and displays an increasing trend, this model was chosen. We used the AutoARIMA model because the model itself will generate the optimal parameter values which would be suitable for the data set to provide better forecasting [15].

7.1.4 *Holt's Linear Trend Model :*

Holt's Model allows you to predict time series data that has a pattern and/or seasonality by extending the simple exponential smoothing process [16]. There are three distinct equations in this model that function together to create a final prediction. The first is a simple smoothing equation which adjusts the last smoothed value directly to the trend of the last time. In the second equation, the trend itself is modified over time where the trend is measured as the difference between the last two normalized values. Finally, to produce the final prediction, the third equation is used. Two parameters are used by Holt's algorithm, one for the absolute smoothing and the other for the pattern smoothing equation [18]. In order to reduce the weight value for the older data, Holt's smoothing allocates exponentially diminishing weights and values against historical data. In other words, more modern historical evidence is assigned to predict more weight than the older outcome [17].

7.2 Evaluation

From Table 1 we observe that all models do generally well in forecasting the future values. On average for all sectors, the Simple Exponential Smoothing model performs the best, with least errors on all three metrics as seen in Table 2.
We then use the SES model to forecast the employment ratios for all three sectors for the coming five years i.e. from 2021 to 2025. These values are then used to calculate the corresponding GEPI values for each year. We analyze the average of the predicted GEPI values for the years 2021 to 2025 with the average of the existing GEPI values for the years 2016 to 2020 based on different geographical locations and income groups.

7.2.1 *Analysis of Predicted GEPI based on the Regions :*

When we take a look at the Figure 17 we observe that unfortunately majority of the region's GEPI values decrease over the next five years (2021-2025). Looking at European & Central Asian regions and Middle East & North African regions around 65% of the countries and 71% of the countries respectively will have their GEPI decrease. This indicates that the these regions need to start planning and implementing effective measure to promote gender equality in the workplace. On the positive side, for the South Asian and Sub Saharan African regions we observe that around 75% of the countries and 68% of the countries respectively will see an increase in their GEPI. The countries that have their GEPI decrease should look at the policies and laws implemented by other countries that have their GEPI increase within the same region. Since the countries are located within the same region, implementing the laws should be easier as compared to adopting laws from a completely different region.

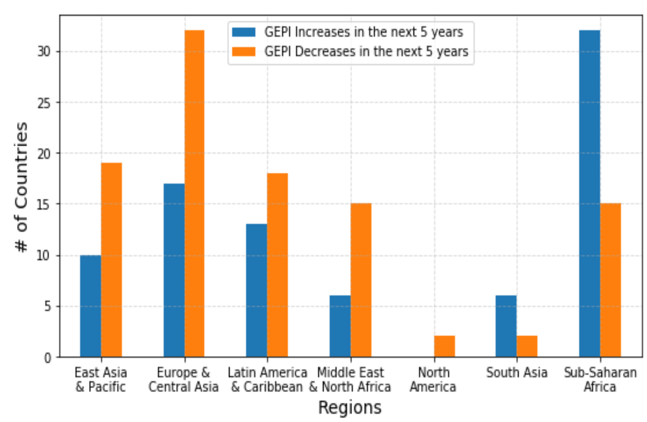

Figure 17: GEPI Analysis based on Different Regions

7.2.2 *Analysis of Predicted GEPI based on Income Groups :*

Looking at the Figure 18 we observe surprising results. A majority of the countries belonging to the High Income and Upper Middle

Income groups will see their GEPI decrease in the next five years. On the other hand, a majority of the countries in the Low Income and Lower Middle Income bracket notice their GEPI increase over the next few years. 75% of the High Income group countries GEPI decreases. These forecasts further strengthen the earlier discussion about the higher gender bias in High income countries compared to low income countries.

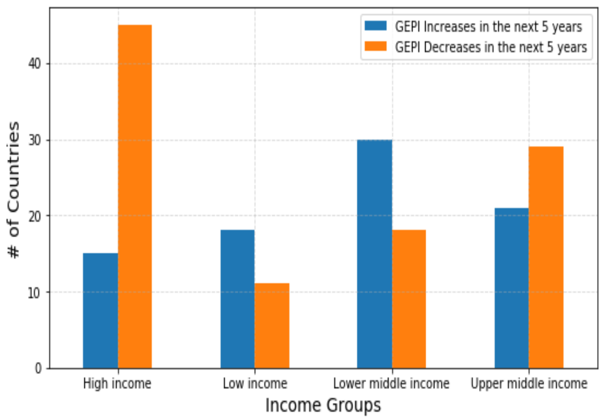

Figure 18: GEPI Analysis based on Different Income Groups

8 DISCUSSION

We'd like to thank the instructor and the support staff of this course, for providing us with this opportunity to do valuable research. This project enabled us to gain perspective on handling large amounts of data as well as missing or incomplete values. We achieved a deeper understanding on the significance of data visualization and various tools and libraries for the same. We also learned the significance of fetching good data sets and conducting thorough research. In future, this project can be enhanced by gathering a more complete and comprehensive literacy dataset which could point to a stronger correlation between employment rate and literacy rate. This data could also help extend our research to performing causality analysis between education and employment as well as gender bias across manual and mental labor. We would also like to research more about the employment situation in different regions. For example, even if a country shows a high female working population, do they get equal pay and have job security as compared to their male colleagues? In the late 20^{th} century, women made more rapid advances in the private sector than they did in the political world. Hence, this project can be extended to analyze the correlation between women in country's leadership positions.

9 CONCLUSION

Gender discrimination is observed in all facets of life. The employment sector in particular suffers from large gender disparities. Our research explores these biases in the employment sector for various countries. We successfully define a Gender Employment Parity Index to recognize any gender bias in employment opportunities. Our results display on average an increasing trend towards gender equality in the workplace. We also analyze the GEPI for different geographical regions and income groups to detect interesting patterns. We observed high proportion of females working in the Sub Saharan regions but also realized most of these women don't have equal pay or job security. High income countries have lower percentage of female workers as compared to low income countries. By analyzing the contributing family worker data, we observed that the gender bias in family worker is still increasing over the years. Next, we examine the correlation between education and employment. The employment rate has different reactions to the change of the literacy rate depending on the industry sector type. Finally we were successful in designing four time series forecasting models to estimate GEPI for the next five years. The predicted GEPI values help prepare the countries next steps in formulating legislature to promote gender equality in the workforce.

REFERENCES

[1] [n.d.]. Labor Social Protection Indicators.
[2] Mariangela Zenga Aglaia Kalamatianou Adele H. Marshall. 2020. Academic Students' Progress Indicators and Gender Gaps Based on Survival Analysis and Data Mining Frameworks. In *Social Indicators Research volume 151, pages1097–1128(2020)*.
[3] Jody Heymann Adva Gadoth. 2020. Gender parity at scale: Examining correlations of country-level female participation in education and work with measures of men's and women's survival. In *Gender parity at scale*.
[4] Revathi Balakrishnan. 2005. Rural women and food security in Asia and the Pacific: Prospects and paradoxes. *Rap Publication* 30 (2005).
[5] Shivangi Chopra, Abeer Khan, Melicaalsadat Mirsafian, and Lukasz Golab. 2019. Gender Differences in Science and Engineering: A Data Mining Approach. In *EDBT/ICDT Workshops*.
[6] Gaëlle Ferrant and Alexandre Kolev. 2016. The economic cost of gender-based discrimination in social institutions. *Organisation for Economic Co-operation and Development Center* (2016).
[7] Darcy Hango. 2013. Gender Differences in science, technology, engineering,mathematics, and computer science (STEM) programs at university. In *Insightson Canadian Society*.
[8] Nathanael A Heckert, James J Filliben, C M Croarkin, B Hembree, William F Guthrie, P Tobias, and J Prinz. 2002. Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods. (2002).
[9] Rob J Hyndman and George Athanasopoulos. 2018. *Forecasting: principles and practice*. OTexts.
[10] Haiko Lietz and Claudia Wagner Mohsen Jadidi, Fariba Karimi. 2018. Gender disparties in science? Dropout,Productivity,Collaborations and success of male and female computer scientists. In *World Scientific*.
[11] Status of Women in the States. [n.d.]. Gender Differences in Sectors of Employment.
[12] Sara Olk. 2018. The State of Female Employment in Sub-Saharan Africa. Blog. https://www.borgenmagazine.com/the-state-of-female-employment-in-sub-saharan-africa/
[13] Ana Petrovska, Patricia Goldberg, Anne Brüggemann-Klein, and Anne Nyokabi. 2020. Mining Gender Bias: A Preliminary Study on Implicit Biases and Gender Identity in the Department of Computer Science at the Technical University of Munich. In *Software Architecture*, Henry Muccini, Paris Avgeriou, Barbora Buhnova, Javier Camara, Mauro Caporuscio, Mirco Franzago, Anne Koziolek, Patrizia Scandurra, Catia Trubiani, Danny Weyns, and Uwe Zdun (Eds.). Springer International Publishing, Cham, 138–150.
[14] Selva Prabhakaran. 2020. ARIMA Model – Complete Guide to Time Series Forecasting in Python. Blog. https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/
[15] Sushmitha Pulagam. 2020. Time Series forecasting using Auto ARIMA in python. Blog. https://towardsdatascience.com/time-series-forecasting-using-auto-arima-in-python-bb83e49210cd
[16] Vera Shao. 2020. Forecasting with a Time Series Model using Python: Part Two. Blog. https://www.bounteous.com/insights/2020/09/15/forecasting-time-series-model-using-python-part-two/
[17] Smarten. 2018. What is the Holt-Winters Forecasting Algorithm and How Can it be Used for Enterprise Analysis? Blog. https://www.elegantjbi.com/blog/what-is-the-holt-winters-forecasting-algorithm-and-how-can-it-be-used-for-enterprise-analysis.htm
[18] P. M. Swamidass (Ed.). 2000. *Forecasting models, Holt'sHOLT'S FORECASTING MODEL*. Springer US, Boston, MA, 274–274. https://doi.org/10.1007/1-4020-0612-8_409

[19] Yves Gingras Blaise Cronin Cassidy R. Sugimotoo Vincent Larivière, Chaoqun Ni. 2013. Bibliometrics: Global gender disparities in science. In *NATURE*.

10 APPENDIX

10.1 Honor Code Pledge

University of Colorado, Boulder Honor Code Pledge : On my honor, as a University of Colorado Boulder student, I have neither given nor received unauthorized assistance."

10.2 Individual Contribution

(1) **Srihaasa Pidikiti**
- Data collection and pre-processing of the chosen employment features from WorldBank data sets (Section 3).
- Calculation of the Gender Employment Parity index. Analysed the GEPI differences with respect to different income group countries. (Sections 4.1, 4.2.1, 4.2.3)
- Analysis of gender bias in different sectors of employment based on the Income Groups (Sections 5.2.4).
- Built the Simple Exponential Smoothing and Double Exponential Smoothing models to forecast female to male employment ratios and in different sectors and GEPI in the future years (2021-2025) and generated the corresponding model evaluation metrics (Sections 7.1, 7.1.1, 7.1.2, 7.2, 7.2.2).
- Worked on the Abstract, Introduction and Conclusions of the report (Sections 1, 9)

(2) **Amatullah Sethjiwala**
- Data collection and pre-processing of the chosen employment features from WorldBank data sets (Section 3).
- Calculation of the Gender Employment Parity index. Analysis of GEPI based on the different geographical locations. (Sections 4.1, 4.2.2)
- Analysis of Gender Bias in the various factors of employment (Sections 5.1, 5.2.1, 5.2.2, 5.2.3)
- Built the ARIMA and Holt's Linear Trend models to forecast female to male employment ratios and in different sectors and GEPI in the future years (2021-2025) and generated the corresponding model evaluation metrics (Sections 7.1, 7.1.3, 7.1.4, 7.2, 7.2.1, 7.2.2).
- Worked on the Abstract, Introduction, Related Works and Conclusion of the report (Sections 1, 2 & 9)

(3) **Xuefei Sun**
- Data pre-processing and cleaning of some datasets with missing date, for example, the literacy dataset (Section 3).
- Analysis of Gender Bias in the contributing family workers for both male and female (Section 5.2.5).
- Computation of datapoints used for correlation analysis of employment rate & the literacy rate and GPI & the GEPI (Section 6.1).
- Analysis and comparison of the correlation between employment rate and literacy rate based on genders and six different regions (Section 6.2.1).
- Worked on the Related Works and Conclusion of the report (Section 2 & 9)

(4) **Preethi Vijai**
- Data pre-processing and data cleanup of GPI vs Gender Employment data on a region basis.
- Calculated correlation between GPI and GEPI in different regions and plotted their correlation using linear regression as well as calculated their corresponding r-values(Section 6.2.2).
- Validated the above calculation further using Pearsson's Coefficient (Fig 16)
- Worked on the Related Works, Discussion and Future Works section of the report (Section 2 & 8)