

Data Analyzing Immigration to Canada using Predictive Analysis (Multiple Linear and Non-Linear Regression)



P. Deeraj, K. Hari Kiran, M. Hemanth Varma, J. Siva Priya

Abstract: *The immigration to Canada impacts the government in different manner like increase in population, waste, fossil fuel and it also benefits like increase economic growth, trade which will increase the GDP value of Canada, increase in workforce of country, open market, globalization, technologies and adapt to different cultures, food, and people^[1]. These would result in a decrease in discrimination and aware about their rights and duties. The immigrants are more interested in entrepreneurship than others^[2]. Which would impact increase in development in the country. The work explores the impact of immigration to Canada from all around the world. The top 5 countries that immigrate to Canada is analyzed by using Jupyter notebook. The prediction is done only for the top 5 countries that immigrate to Canada by analyzing the previous immigrants from 1980 - 2013. The multiple linear regression is used to analyze the data.*

Keywords : *Entrepreneurship, Multiple Linear Regression, Globalization, Workforce.*

I. INTRODUCTION

The intention of this research is to analyze the data and visualized the immigrants to Canada from top 5 countries (India, China, United Kingdom, Philippines and Pakistan). The data is analyzed by Jupyter notebook and creating of feature set. The data is reshaped and reworked to obtain the data as requested to do further changes. Plotting of graph for the model obtained by analyzing the data using multiple linear or multiple nonlinear regression it can be performed by numerous ways like by using Azure Machine Learning Studio or by Jupyter notebook.

II. DATA WRANGLING & SCRUTINIZE

The data gathered for Canada immigration from all 197 countries is in the form of raw data, the data is filled in an excel file and that is converted to an .csv file. The .csv file is uploaded in Jupyter notebook by using NumPy, Pandas, matplotlib library files. The is modify to be very effective to use.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

P. Deeraj, Department of Computer Science Engineering SRM Institute of Science and Technology, Chennai, India

K. Hari Kiran, Department of Computer Science Engineering, SRM Institute of Science and Technology, Chennai, India

M. Hemanth Varma, Department of Computer Science Engineering, SRM Institute of Science and Technology, Chennai, India

J. Siva Priya, B. E, M.E, Assistant Professor, Department of Computer Science Engineering, SRM Institute of Science and Technology, Chennai, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

A. Source of Data

1. [United Nation department of economic and social affairs.](#)

Dataset features:

1. Type
2. Coverage
3. Country name
4. Area
5. Area name
6. Region
7. Region name
8. Dev
9. Dev name
10. 1980 - 2013 Years

Years holds the number of immigrants to Canada from each country.

Structure: The raw data is in the form of text document with comma (',') as separator between each attribute and the new entry record is separated by new line. The raw data is converted to .csv by using excel with comma (',') as separator with the attributes. The .csv file is uploaded to Jupyter notebook by using library file like NumPy, pandas and matplotlib. The top 5 countries that immigrate to Canada are found using multiple linear regression.

[World bank open data](#)

Dataset features:

1. Population growth
2. Education growth
3. Income
4. CO2 damage
5. Natural resources

These data are taken for india, china, united kingdom, philippines and Pakistan because they are the top 5 countries immigrate to Canada by the analysis made using jupyter notebook.

Structure: The is directly used to do multiple linear regression because the data is already in the form of .csv file. The file is combined with the canada immigration data to know the number of immigrants.

B. Top 5 Countries

Country	India	China	United Kingdom of Great Britain and Northern Ireland	Philippines	Pakistan	
1980	8880	5123		22045	6051	978
1981	8670	6682		24796	5921	972
1982	8147	3308		20620	5249	1201
1983	7338	1863		10015	4562	900
1984	5704	1527		10170	3801	668

Fig. 1. Top 5 countries

III. REGRESSION

There is more than one technique to do this operation. The data can be branched into two groups 1. Unsupervised technique, 2. Supervised technique. The Unsupervised technique is used for clustering, dimension reduction, density estimation, market basket analysis. The data in unsupervised are all unlabeled and the model works by its own to unrevealed new information. The supervised technique is used for classification and regression models, there are used to predict the future data by analysis the vast amount of previous data. The work uses regression model to predict the data instead of classification model because in our data it is not classifying anything like whether the developing countries or developed countries are immigrating. The work focuses only on predicting the increase number of immigrants to canada for top 5 countries.

The model of multi-variable linear regression can be represented by the following:

$$Y = a_1 * X_1 + a_2 * X_2 + a_3 * X_3 + \dots + a_n * X_n + b$$

The work used the multiple regression model because it is using multiple independent variable to predict one dependent variable

IV. MULTIPLE REGRESSION

The multiple regressions can be done easily using microsoft excel sheets instead of using jupyter notebook. The data of the top 5 countries that immigrated to canada from 1980 - 2013. The data separate the countries into 5 different feature sets and for each country adding 6 new parameters.

For all top 5 country feature set:

1. Years [1980 - 2013]
2. Population growth
3. Education growth
4. Income
5. Nature expense
6. CO2 damages
7. Number of immigrants to Canada

V. EXCEL OPERATION

The work is using excel to do multiple regression both linear and nonlinear models. By the model it is able to find the accuracy of the model to predict for future data by using the independent variables. Before doing multiple regression in excel the work need to check it have the necessary data

analysis toolkit in the file. It can be easily done by clicking file on the menu bar and select 'option' then 'add-ins' and click on the check box of analysis toolkit.

1. X - independent variables (can be one or more variables)
Includes: Year, Population growth, Education growth, Income, CO2 damage, Natural resources
2. Y - dependent variable (the one predicts variable)
Includes: Number of immigrants to Canada

The x represented as variables that are used to find the y variable. The is going to predict the number of immigrants and its accuracy for the top 5 countries

VI. METHODOLOGY

The number of immigrants to canada from each country will be analyzed and will provide us the accuracy and error of the model. If the error is more the model will be useless to predict the number of immigrants to canada from that country.

Error > 5000, The standard error is the average standard variance of the population, So the standard error should be less if there are more 5000 variances between the points, then we assume it will have high error. Then the model is not useful, if the error is more than 5000 with that huge difference the work can't predict the number of immigrants to canada with high accuracy.

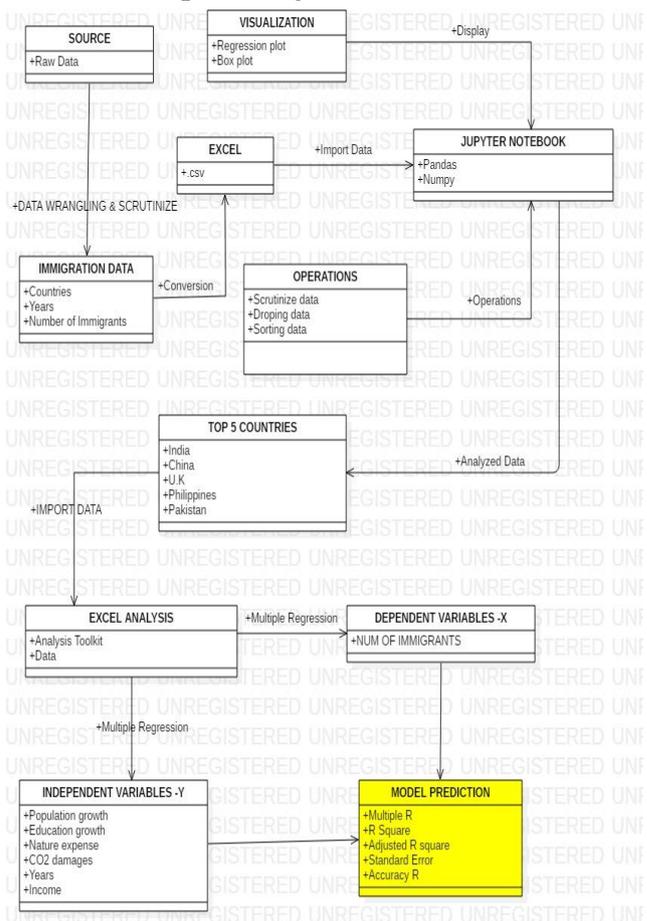


Fig. 2. Architecture Diagram

VII. RESULT AND DISCUSSION

The work analyses the population growth, education growth, co2 damage, income, years and natural resources are the independent variables, which are used to predict the dependent variable number of immigrants to canada from the top 5 countries, they are india, china, united kingdom, philippines and pakistan.

A. India – Canada Immigration

India is a developing country with vast population. There are 1.1 billion people and it will increase in the upcoming years. The huge growth of population has led to be the main reason for increase in immigrants. The prediction is made for number of immigrants to canada from india. The prediction results as can out it looks very good.

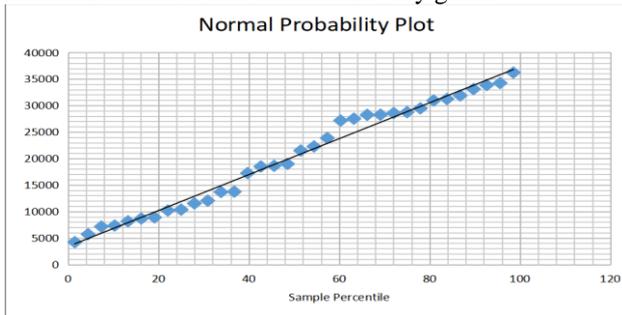


Fig. 3. Percentile vs Percentile india immigration

1. Accuracy (r) - 94.8%
2. Error - 3175 (less)

Table I. India Statistic output

Regression Statistics	
Multiple R	0.9579341594
R Square	0.9176378537
Adjusted R Square	0.8993351545
Standard Error	3175.102009
Observations	34

B. China – Canada Immigration

China is the most populated country in the whole world. It has over 1.3 billion people with 5 time zones. China is growing 0.5% of population per year. By this growth rates the number of youngsters also increases with great skills and intelligence. China is becoming more powerful day by day on military, developments and technology. Asian’s are becoming more influential for future technologies. These made other countries to hire or work with them because of their hardworking skills and intelligence. These job offers from multinational companies are increasing and getting a greater paycheck.

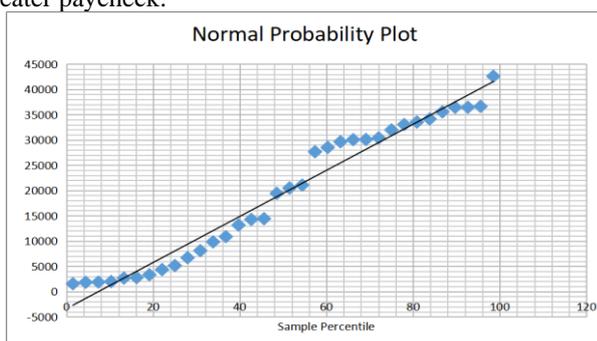


Fig. 4. Percentile vs Percentile china immigration

1. Accuracy (r) – 96.3%
2. Error - 3634 (less)

Table II. China Statistic output

Regression Statistics	
Multiple R	0.9701973613
R Square	0.9412829198
Adjusted R Square	0.9282346798
Standard Error	3634.800136
Observations	34

C. United Kingdom – Canada Immigration

United Kingdom is a developed country where there was a time when people immigrated to canada. The immigrants to canada from the great britain were so high that was even named “Great Migration of Canada”. At that time over 800,000 people were immigrated to canada from europe. At present the immigrants to canada is decreasing each year.

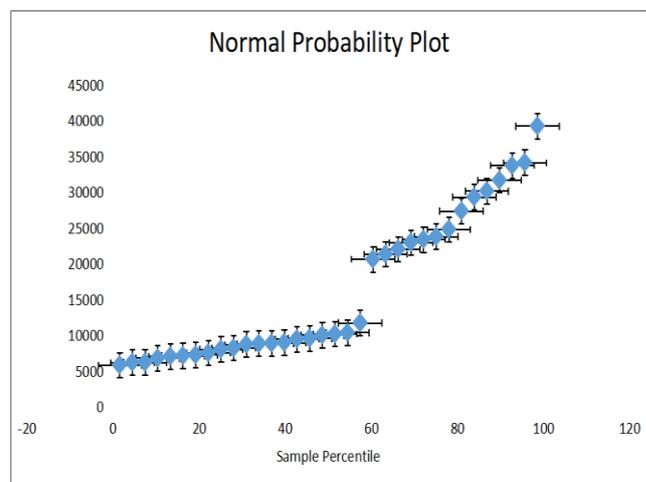


Fig. 5. Percentile vs Percentile united kingdom immigration

1. Accuracy (r) – 72.02%
2. Error - 7122 (high)

Table III. United Kingdom Statistic output

Regression Statistics	
Multiple R	0.7786403293
R Square	0.6062807624
Adjusted R Square	0.5187875985
Standard Error	7122.666829
Observations	34

D. Philippines – Canada Immigration

Philippines is overpopulated country with a small amount of surface area. Philippines has become the 3rd biggest groups in canada. There was a tremendous increase in immigration to canada from 1970’s due to instability of economic and difficulties in government. So, people tended to immigrate to canada and it will continue to increase in the future too.

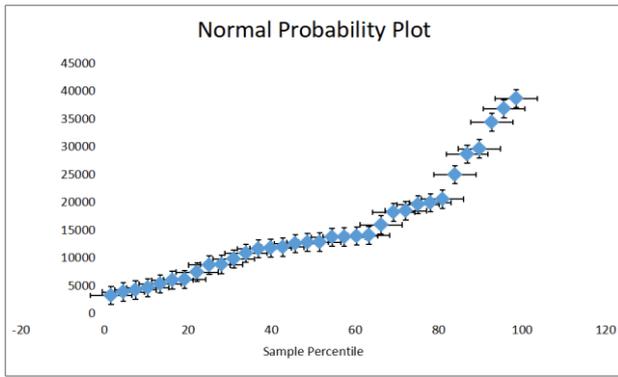


Fig. 6. Percentile vs Percentile philippines immigration

1. Accuracy (r) – 91.01%
2. Error – 3938 (less)

Table IV. Philippines Statistic output

Regression Statistics	
Multiple R	0.9271209821
R Square	0.8595533154
Adjusted R Square	0.8283429411
Standard Error	3938.792774
Observations	34

E. Pakistan – Canada Immigration

Pakistan is the world's 6th most populated country. The country faces major economic crisis both in finance and development. The increase in national debt as made the country to fall apart. This made people to immigrate to canada to find better life and good health care system

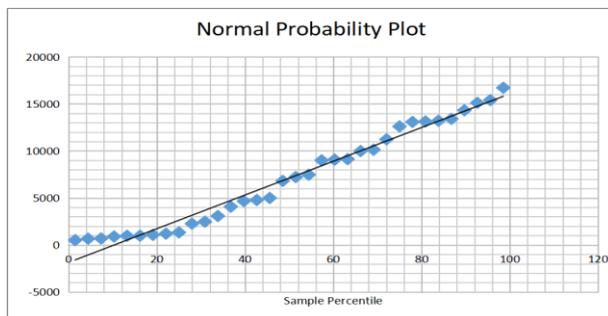


Fig. 7. Percentile vs Percentile pakistan immigration

1. Accuracy (r) – 97.3%
2. Error - 2583 (less)

Table V. Pakistan Statistic output

Regression Statistics	
Multiple R	0.8982212187
R Square	0.8068013577
Adjusted R Square	0.763868326
Standard Error	2583.151078
Observations	34

VIII. CONCLUSION

The immigration to canada from the top 5 countries has been analyzed and found that the immigration from india is

going to be increasing with an accuracy of 94.8%. The immigration from China will be increasing with accuracy of 96.3%, The immigration from the United Kingdom will be increasing with less accuracy of 72.02%, The immigration from philippines will increase with accuracy of 91.01%, The immigration from pakistan will also increase with accuracy of 87.3%. The accuracy is found using Multiple regression and the top 5 countries are found using jupyter notebook with operations like NumPy, pandas and matplotlib. The growth of immigrants to canada will result in an increase in the workforce, technology, economic growth, open markets, globalization and it is a win- win for both the people and the country.

REFERENCES

1. Canada citizenship and immigration resource center
2. Harvard Business Review
3. International Migration: A Panel Data Analysis Of Economic and Non-Economic determinants by Anna Maria Mayda on May 2005 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=72544
4. Student Flow and Migration: An Empirical Analysis by Axel Dreher and Panu Poutvaara on March 2006 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=731765
5. Analysis of the Entrepreneurial Immigrant Profile in Spain by Genoveva Millán, Virginia Navajas and Ricardo Hernández on April2018 https://www.researchgate.net/publication/330881048_Analysis_of_the_Entrepreneurial_Immigrant_Profile_in_Spain
6. Quality of Work Experience and Economic Development: Estimates Using Canadian Immigrant Data by Serge Coulombe, Gilles Grenier, Serge Nadeau on January2011 https://www.researchgate.net/publication/254439974_Quality_of_Work_Experience_and_Economic_Development_Estimates_Using_Canadian_Immigrant_Data

AUTHORS PROFILE



P. Deeraj, Currently studying 3rd Year Bachelor of Computer Science Engineering, batch (2017 - 2021) from SRM Institute of Science and Technology (formerly known as SRM University), Ramapuram, Chennai. Internship Experience: Python at UNIQ technologies, T Nager, Chennai. Area of interests: Data Analysis, Predictive Analysis, Cloud Computing. Online Certificates: IBM PY0101EN Python Basics for data science, IBM DA0101EN Analyzing data with python, IBM DV0101EN Visualizing data with python, IBM ML0101EN Machine learning with python, Microsoft DEV276x Learn to program in java, Microsoft DAT229x Introduction to Professional oriented Big Data, Microsoft DAT206x Analyzing and visualizing data with excel. Projects: Twitter Bot, Mac changer. Current / ongoing project: Creating a tech info channel website, podcast and YouTube channel. Future goals: Entrepreneurship. Social media LinkedIn profile: <https://www.linkedin.com/in/deerajpannam/> Email: deeraj.pannam@gmail.com



K. Hari Kiran, Currently studying 3rd Year Bachelor of Computer Science Engineering, batch (2017 - 2021) from SRM Institute of Science and Technology (formerly known as SRM University), Ramapuram, Chennai. Internship Experience: API testing at India infotech, MMDA, Chennai. Area of interests: cloud privacy and security. Projects: Twitter Bot, Mac changer, Climate master php page. Current / ongoing projects: Creating a tech info channel website, podcast and YouTube channel. Future goals: Network Security Engineer or Cyber Security Analyst. Other Activities: Member of NSS, Volunteer for social activities and club events like Beach cleaning and Blood donation camps. Social media LinkedIn profile: <https://www.linkedin.com/in/hari-kiran-002> GitHub: <https://github.com/A2hari/> Email: harikiranavarthapu@gmail.com



M. Hemanth Varma, Currently studying 3rd Year Bachelor of Computer Science Engineering, batch (2017 - 2021) from SRM Institute of Science and Technology (formerly known as SRM University), Ramapuram, Chennai. Internship Experience: Python at UNIQ technologies, T Nager, Chennai. Area of interests: IOT (Internet of Things), Software Testing. Projects: Twitter Bot, Mac changer, Climate master php page. Current / ongoing projects: Creating a tech info channel website, podcast and YouTube channel. Other Activities: Member of NSS, Volunteer for social activities and club events like Beach cleaning and Blood donation camps Future goals: Software Development Engineer or Web Development Engineer.

Twitter: Varma1617

Email: varmahemanth16@gmail.com



J. Siva Priya, B.E M.E, Assistant Professor (O.G.), Area: DBMS, Big Data, Machine Learning. Affiliation: Department of Computer Science and Engineering, Ramapuram Campus, SRM Institute of Science and Technology (formerly known as SRM University). M.E. CSE at SSN college of engineering.

Course Subject: Software Engineering, Object Oriented Analysis and Design. Research Interests: Big Data and Machine learning. Academic Experience: Working as an Assistant Professor in Department of CSE, SRMIST, Ramapuram from June 2018 to till date. Professional Experience: Worked as Assistant Software Engineer in TCS for 1 year, Worked as Engineer in Honeywell technology solutions, Madurai for 1 year. Achievements and Awards: Awarded as the best team member in Honeywell technology solution in the year 2017. Email: sivaprij@srm.ist.edu.in