

WORLDWIDE LABOUR MIGRATION ANALYSIS USING LINKEDIN DATA



by-
Sudhish Subramaniam
Urvashi Dube
Dishant Gupta
Dhaval Jariwala
Yesha Gosaliya



TABLE OF CONTENTS

- 01** PROBLEM STATEMENT
- 02** UNDERSTANDING THE DATA
- 03** DATA MODELLING
- 04** DEMO AND RESULTS
- 05** REFERENCES
- 06** CONCLUSION





02 PROBLEM STATEMENT- WHAT, WHY, HOW?

WHAT IS THE PROBLEM?



In today's world, people migrate from one place to the other for growth and development.

The most common purpose of migration is a money hike. According to statistics, not every person who migrates has profited. When people leave their homeland they feel stuck, lonely, and depressed, and face problems such as lack of opportunities, coping with a different common language, homesickness, lack of family support, and cultural differences in the migrated country.

All these problems lead to underperformance in their career and ultimately not completing their goal of a money hike. People are always concerned if their decision of migration would be profitable or not.

WHY IS IT WORTH FURTHER RESEARCH?

- It is necessary to solve this issue, because it decides the career of a person.
- The future of a person totally depends on the target country the person is migrating into.
- Companies in the target country can use the model to analyse people coming from the base country.
- The analysis can also be used to examine the economy of countries.



HOW CAN WE SOLVE THE PROBLEM?



This problem can be easily solved by regression analysis.

Our project has developed an Rshiny app to visualize the base country's and target country's advantages based on country, industry and skill migrations. The prospective migrant can choose base country and their target country with the preferred industry/ skill to check if their would be growth or downfall.

We have also developed a regression model based on Random Forest to predict absolute net flow of money for the year 2023. This will also help people to understand the growth in the future.

02

UNDERSTANDING THE DATA



OUR DATASET



Dataset link:

[https://datacatalog.worldbank.org/search/dataset/0038044/Talent-Migration---LinkedIn-Data-.](https://datacatalog.worldbank.org/search/dataset/0038044/Talent-Migration---LinkedIn-Data-)

The dataset is extracted from LinkedIn's Economic Graph Initiative and World Bank Group, and is available at country, country-industry and country skill level based migrations.

The migration in the dataset is recorded when a LinkedIn member changes location on their LinkedIn profile.

The 3 datasets-

01 Country Migration

02 Industry Migration

03 Skill Migration

OUR DATASET- IN DEPTH

The types of datasets used -

01 Country Migration

02 Industry Migration

03 Skill Migration

Important columns used in the dataset-

Country names (both base and target countries) with their latitudes and longitudes(for world map), Regions, Industry name, Skill group name, Skill name and Net flow of money every year were taken into consideration.

Method used in analysis?

Country migration and industry migration data were merged by countries and similarly country migration and skill migration data were merged by countries to get more accurate analysis and predictions.



03 DATA MODELLING



HOW DID WE MODEL OUR DATASET?

We have used the net flow of money in the 2015,2016, 2017, 2018 to train our model. We first predicted the netflow of money for each country for the year 2019. This helped us to build a model to predict the netflow of money for the year 2023. We do not consider the year 2020, 2021, 2022, because of the COVID 19 Pandemic. This is because during these years the labour migration was next to negligible in all the countries. Hence, we predict for the year 2023.

Two key implementations from our app:

- To help a person determine to which part of the world he/she should migrate in the upcoming year 2023, we have compared the net flow of money for the year 2019 and the predicted netflow of money for 2023. This would be helpful for prospective migrants to understand the heightening or unseating of an area.
- We have predicted the absolute net flow of money for the year 2023 for each country, based on country, industry and skills separately. This can be used as the key factor to whether or not to migrate to a specific country.

04

MODEL DEMO

RESULTS

Hence, we predict for the year 2023. On applying Linear Regression and Random Forest for our predictions, we found that R2 Score of Random forest is 0.9614647 and R2 score of Linear regression model is 0.9235135, which clearly indicates that Random forest regression is a better fit model for this dataset.

```
Call:  
lm(formula = net_per_10K_2019 ~ net_per_10K_2015 + net_per_10K_2016 +  
    net_per_10K_2017 + net_per_10K_2018, data = country_wise_net_per_year_lr)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-21.510 -3.907 -0.546  5.451 40.886  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -1.26986  0.94750 -1.340  0.1824  
net_per_10K_2015  0.10147  0.05735  1.769  0.0791 .  
net_per_10K_2016 -0.01945  0.09487 -0.205  0.8379  
net_per_10K_2017 -0.24115  0.10918 -2.209  0.0289 *  
net_per_10K_2018  1.06370  0.07639 13.925 <2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 9.473 on 135 degrees of freedom  
Multiple R-squared:  0.9235,   Adjusted R-squared:  0.9212  
F-statistic: 407.5 on 4 and 135 DF,  p-value: < 2.2e-16  
  
[1] "RMSE Linear Regression"  
[1] 9.302516  
[1] "R2 Linear Regression"  
[1] 0.9235135
```

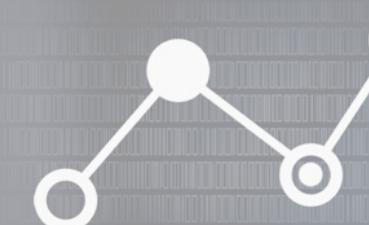
Linear Regression Summary, RMSE,R2,

year_2019_rf	test_2019_rf
Min. : -132.1600	Min. : -96.837
1st Qu. : -18.0700	1st Qu. : -15.824
Median : -3.8400	Median : -3.966
Mean : -0.6739	Mean : -1.020
3rd Qu. : 10.5500	3rd Qu. : 11.213
Max. : 221.4300	Max. : 158.850

```
[1] "RMSE Random Forest"  
[1] 8.285805  
[1] "R2 Random Forest"  
[1] 0.9614647
```

Random Forest Summary, RMSE,R2,

05 CONCLUSION





WHAT DID WE CONCLUDE?

- We have successfully developed a Rshiny app by deploying our machine learning model to predict the netflow of money for the year 2023 for each country and each region.
- The exploratory data analysis shown in the Rshiny app would be useful to any prospective migrant who is deciding to which country he/she is migrating to.
- The data analysis and predictions can be used by companies to understand the economies of other countries and regions. This can help them expand their business in different potential countries and regions in the upcoming years.
- The model can be used to specifically to analysis the growth or downfall for each country, industry and skill.
- The interactive world maps in the Rshiny app helps to visualize countries and regions development more deeply.

REFERENCES

- <https://r-graph-gallery.com/>
- <https://shiny.rstudio.com/>
- <https://www.stt.msu.edu/~melfi/STT180Text/shiny-interactive-web-apps-in-r.html>
- <https://sesync-ci.github.io/basic-Shiny-lesson/index.html#:~:text=Shiny%20is%20a%20web%20application,share%20results%2C%20and%20much%20more>
- <https://datacatalog.worldbank.org/>
- <https://www.r-bloggers.com/2021/04/random-forest-in-r/#:~:text=Random%20Forest%20in%20R%2C%20Random,to%20Identify%20the%20important%20attributes>





THANKYOU

by-

Sudhish Subramaniam
Urvashi Dube
Dishant Gupta
Dhaval Jariwala
Yesha Gosaliya