SAPIENZA
UNIVERSITÀ DI ROMA

Performance of Knowledge Organisation and Human Action using Probability Graphical Model and Efficiency Analysis

Group: F
Trina Sahoo (1901254)
Debodeep Banerjee(1901253)
Octavianus Sinaga(1907453)
Ulviyaa Jafarli(1921169)

Contents

1 Abstract

The purpose of this study is to do systematic review on the performance of knowledge organizations based on modelling human action, assessing quality and intellectual capital. We have focused in the relationship of modelling human action and the performance of knowledge organization. Articles were obtained from Scopus and Web of Science (WOS). A total of 71 articles are obtained avoiding the duplicates from Scopus and WOS. These articles are reviewed, analyzed and important information was extracted. The keyword frequency, documents relevance and relevant field of subjects was obtained on these articles. The findings indicates that majority of the articles were extracted from Scopus and the subject area covered are Business Management Science, Computer science, Social science and Decision science. A reviewed studies were conducted on the mentioned subjects including Economics, mathematics etc. Findings were discussed briefly and it was recommended for further studies of the relevant paper which contains the highest frequency keywords such as knowledge management, intellectual capital, knowledge organizations and human capital. The whole structure can also be performed for a different set of papers which involved different query search and also we can use more visualizations. As the focus was mainly based on human action, the ontological models have been shown to provide a graphical view. Moreover, Probability graphical model and efficiency analysis have been performed to observed the imminent role of knowledge organization and human action in a vivid manner.

2 Introduction

Our main aim is to go deeper into the term knowledge organisation. Now, before going to anywhere, the question that hits us is, "what is knowledge organisation?". The term knowledge organisation essentially have two dimensions. While one dimension plays with an individual person, the other focuses more on the public aspect [1]. Knowledge organisation stands for an internal process of transforming or collection knowledge at a person's individual level. On the other hand, in macro level, knowledge organisation stands for gathering information for the public usage. Thus it becomes more collaborative and ivolves collection, processing of information and documemts as a chamber of knowledge [2].
Now, when the concept and aspects of knowledge are being discussed, the concept that pops up ievitably is human actions. Human beings and their actions are a prime factor behind the processing and the development of knowledge. But every movement cannot be termed as human action.Human action is necessarily a purposeful activity [3].
So far we have discussed on knowledge organisation and human activity. Human activity plays a major role in the formation of knowledge organisation. Further, we should also discuss on intellectual capital. In any organisation, capital is not always the financial capital. There is intellectual capital as well. Intellectual capital essentially means the entire knowledge acquired by an entire organisation. We can claim that this knowledge organisation, human activity and the intellectual capital can be stitched in a same thread. We know how how knowledge organisation takes place. But in order to form knowledge organisation, we need human action. And eventually, we are able to build a intellectual capital which is nothing but a hub of knowledge acquired by an entire organisation. In this report we try to explore the scope of systematic review under the notion of knowledge organisation, modeling human action and intellectual capital.
Few researchers addressed the link between information technology and knowledge transfer, and even fewer the relationships between motivational leadership and knowledge management. While it is realised in the literature that it might be a little problematic to implement knowledge management as a change agent during organizational change, it is argued that it can be used by what is today referred to as knowledge leaders, to manage not only knowledge creation, storing and codification, but also knowledge sharing to enhance the learning culture of emerging knowledge organizations. The commonality of most recent research indicates an emphasised focus on knowledge management (technical, human and communication components), and it is argued that knowledge leaders should implement strategic integrated communication to ensure the sustainability of knowledge organizations. These knowledge leaders should be able to: empower individuals to respond creatively to changing situations; adopt personal and active attitudes, individual and organizational goals; contribute to resonant leadership practices; should be self- and socially aware of emotions and goals; be equipped with skills such as self- and relationship management which is characterised by transparency, adaptability, collaboration and motivation; and should be associated with a supportive organizational climate due to a constructive organizational culture with the aim to inspire people to learn. Where knowledge management focuses on two main theoretical perspectives, namely human capital and knowledge based theory, the theory of strategic integrated communication emphasises that knowledge leaders should acknowledge the premises of the strategic intent of the

organization by managing information through, inter alia, motivation, innovation and creativity. The research problem is that in spite of the tremendous research opportunities to examine these constructs, limited research has been conducted from emerging knowledge organization perspectives, especially during change and transformation. This study addressed this gap to enhance the field's discussion with the main aim to critically review existing literature based on an interpretivistic approach from a predominantly postmodern perspective. Hence a theoretical framework has been developed to indicate the interrelationship of these concepts which should be considered by emerging knowledge organizations in future to ensure sustainable stakeholder relationships [4]. We also wants to explore the knowledge organization using data units. In this project the dataset is dependent on human action. The dataset that we have used is IBM Human Resource data. Studies on human resource emphasize the contribution of human and human resource management to organizational performance achievement. Human and organizational capitals are strategic capability and mechanism to create value in knowledge organization. From various library research it has been seen that there is an interactive relationship between human resource system as an element of organizational capital and role behavior as an element of human capital, in strategic human resource management, systemically. Role behavior, then, not only is positioned as an output, but also an input for the desired human resource system, as organizational capital. The result of the study needs to be followed up by empirical results, to validate the proposed interactive relationship between human capital and knowledge organizational. In order to establish the relationship we considered the work life balance of an individual considering different parameters in the organizations by analyzing a dataset based on human resource management. We have performed different functionality of various tools, viz, Probability Graphical model and Efficiency Analysis to get the desired outcome.

3 Literature Review

The project explains the knowledge organizations in a broader aspects and modelling human action, intellectual capital and assessing quality as a part of it. By doing literature review of 7 selected papers we have summarized the concept of knowledge organization and human action which is the motive of the project.

The first paper [5] is based on knowledge management which aims at manipulating knowledge by storing and redistributing corporate information that are acquired from the organization's member. The paper introduce the concept of Virtual enterprise in this context which plays a crucial role as a temporary alliances of enterprises joined together to share resource and skills in order to better response to business opportunities. The paper [6] have a better overview about human capital which essentially has the potential to be key part of the overall firm strategy. In a knowledge-based economy, employees are considered to be a primary component for attaining a competitive advantage. This paper used SPSS to establish the fact that human capital operation is focused on forming, prompting and restriction, collocating and flowing of human capital, and the prompting and restriction of human capital is the main body of the operation.

The paper [7] provides a generic outline on the performance indicators and metrics for knowledge management(KM). The study of this paper based on two hypothesis. The first hypothesis is that successful KM in organizations requires the knowledge the linking of knowledge resources to organizational objectives and the hypothesis is supported by successful KM links knowledge resources to company objectives. The second hypothesis is that a precondition for successful KM is that explicit, quantitative indicators are used and the hypothesis is supported by KM in the case is not based on explicit and quantitative indicators. Another paper [8] that deal with KM investigated the domain of knowledge management (KM) in organizations. Starting from 2014 IEEE started big scaled research to decrease the risks and develop management in different sectors. Main concepts such as KM policies and strategies, KM methods and techniques were shown in the paper with explanations. The purpose of the research is to identify essential factors in the effective knowledge management, make a guide for industries which explains step by step process to create effective KM. Another research paper [9] has been used because of its large information about how KM plays an important role to make organization knowledgeable from management aspects. The article focused on importance of intellectual capital assets such as knowledge resources, business data processing and expert or intelligent systems. Various scheduling methods and features, deductive databases involved in the research for the usage of KM. In conclusion, given research identified expert tool GUESS and its principles as a solution service in terms of described concerns.

The knowledge organization and modelling human action is based on various fields like engineering. Finite Beta-Liouville mixture models have proved to be an effective and powerful knowledge representation and inference engine in several machine learning and data mining applications [10]. The demonstration of the effec-

tiveness of the proposed approach through both synthetic data as well as two challenging real-world applications namely human activities modeling and recognition, and facial expressions recognition. Another article [11] describes the details of NKOS workshop and how it turns traditional knowledge organization systems into various classification formats such as schemes. Including different type of new KOSs design systems have and compared with the old traditional methodologies. In the analyses paper explored the standards, principles, methodologies step by step for web applications. It is a qualified guide for understanding how develop Web services to support usage of KOSs systems and libraries.

A brief summary of our literature review is shown with the help of ontological modelling and PRISMA model. The results are obtained with the help of VOSviewer, Biblioshiny and also with the analysis part of Scopus. Scopus is useful for collecting papers which can be used for literature review and at the same time we can analyze the documents with respect to time, author, types, subject area etc in Scopus. VOSviewer takes a distance-based approach to visualizing bibliometric networks. Any type of bibliometric network can be visualized. Directed networks, for instance networks based on direct citation relations, are treated as undirected. Networks may consist of several thousands of nodes. Because of computational limitations and memory constraints, networks with more than 10,000 nodes tend to be difficult to handle in VOSviewer. Biblioshiny works with the help of R package "bibliometrix". The functionality is almost the same as VOSviewer. It helps in easy use of the main features of bibliometrix [12].

3.1 Ontological Modelling

3.1.1 Eddy

The ontological modelling on eddy is based on a generic overview of knowledge organization and modelling human action. The ontology is reported in figure through the GRAPHOL ontological language, which provides a visual representation of OWL ontologies.
Classes are represented in rectangles and object properties by diamonds. To denote the first (domain) and the second (range) component of a property, a white and a black square were respectively used, connected to the properties they refer to with dashed arrows ending with a small diamond. A black hexagon indicates a disjunctive union of classes. In both cases, each class in the union is connected to these symbols with a dashed edge. Labels associated to the domain and range of properties are used to indicate carnality restrictions (similar to carnality constraints in Entity-Relationship). Solid arrows denote inclusions between classes, both atomic, which correspond to the labeled rectangles, and complex [13].

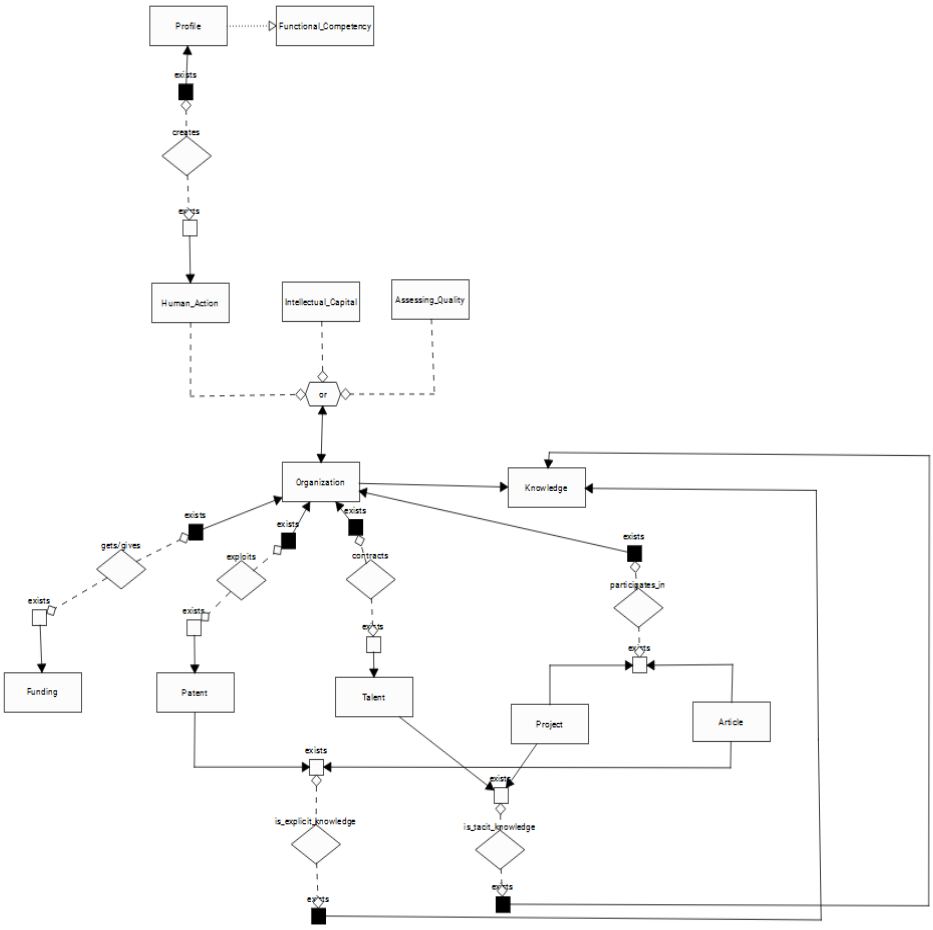


Figure 1: The reference ontology in the case study

It contain 12 classes: *Organization, Knowledge, Funding, Patent, Talent, Project, Article, Human Action, Intellectual Capital, Assessing Quality, Profile and Functional Competency.* A knowledge organization can be human action, intellectual capital and assessing quality. The figure represents these classes along with most relevant properties. The central class in human resource which has further model to profile and functional competency. Functional competencies are expressed in terms of knowledge, skills and abilities. They are also called hard, job related competencies or technical competencies. . All object properties are typed in both the domain and range (indicated by the arrows denoting inclusions from the domain or the range of the property to the class it is typed on), which means that no individuals other the ones among the instances of the typing classes can instantiate the domain and the range of the object property. Similarly, for the domain of data properties (whereas for simplicity we do not indicate the type of the values used for each attribute, that is we do not specify which value-domain the range of an attribute is included in).

3.1.2 Protégé

According to our Human Resource Management dataset, we have decided to explore ontology-based approach by using Protégé software. Given software assist us to build relation graph of HR dataset. Among 35 attributes in our data, 19 main ones were selected and used to build relational graph. Based on the different types of features, class hierarchy contains 4 topics and each have several sub-elements. Building ontological approach with Protégé demonstrate the common representation of HR field and relation of different important attributes.

Figure 2: The ontological modelling using Protégé

3.2 PRISMA Model

PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) model demonstrate the database of studies and records used for systematic review. As a result of written queries in 2 different resources Scopus and WOS resulted with 58 and 27 articles respectively. By removing the 17 duplicated articles, in total we got 68 articles to screen from both sources. After filtering the articles, 8 of them were excluded for the result of unavailability and we included 60 studies for systematic review. In addition, via other methods such as websites and citation searching 9 more reports had been included in database. In conclusion, we obtained 60 studies and 9 more resources for meta-analyses which belong to several different areas. Preferred items used to write introduction, and find results for the research.

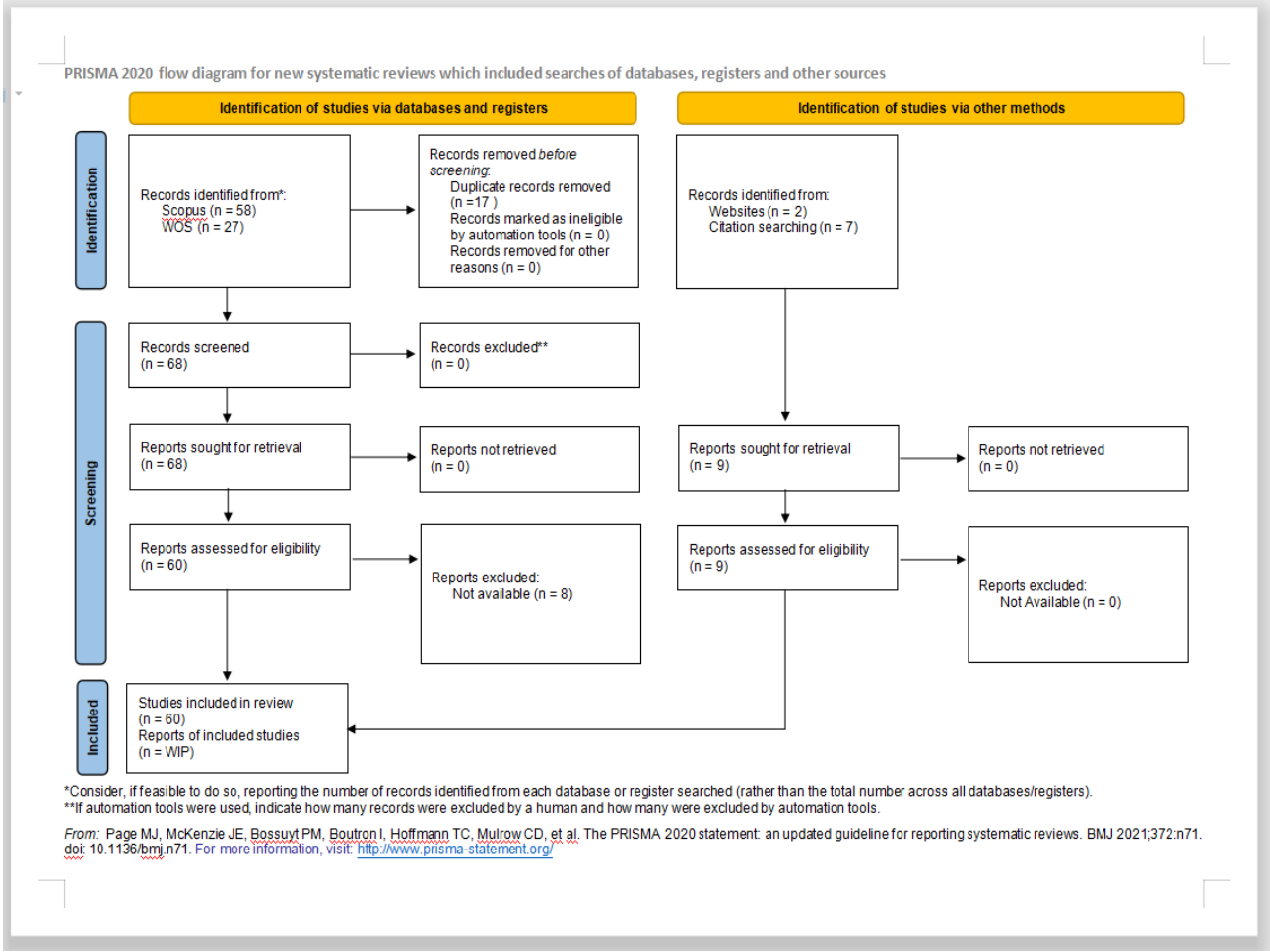


Figure 3: PRISMA model

3.3 Visualization

Result using Scopus

Figure 4 represents the documents according to the subject area. The subject areas like Business Management Science, Computer Science, Social Science and Decision science covers more than 50% of the pie chart which is quite relevant with the topic of our interest. Amongst which Business Management Science covers 30.9% followed by computer science with 16.5% and so on.

[h]

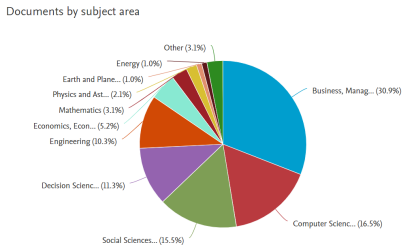


Figure 4: Documents by subject area

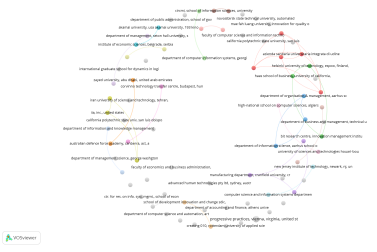


Figure 5: VOSviewer visulazation of organizations

Network visualization of the Bibliographic Data using Organizations and the co-occurrence of Author's Keyword

Figure 5 presents the network visualization of the bibliographic data. The VOS map has been obtained with the connectivity of the organizations. The distance between each cluster represents organizations that are

strongly related in terms of co-authorship. The organizations which are not connected have no connectivity between the publications. There are total of 80 clusters which are fairly small. It will be helpful if the number of clusters can be reduced.

From the above results and graphs we can see that there is a relationship between intellectual capital, human action with knowledge organizations whereas assessing quality is not in the limelight as it is included as a factor in all the factors mentioned. Moreover, the focused subject area is Business Management Science, Social science and Decision Science.

Relevant Words and Clustering by Document using Biblioshiny

The revelant words related to our interest is obtained with the help of Biblioshiny. In the figure 5, it shows that knowledge management, intellectual capital, knowledge organizations and human capital are significant than other. These are the major parts which we are interested in our studies. We are also interested in the effect of assessing quality with knowledge organizations but from the visualizations of both the tools it's quite clear that the effect of assessing quality exists but it is not that much effective as human action and intellectual capital.

Figure 6 represents the clustering by documents. Each clusters represents a bunch of paper which are strongly related by documents. The knowledge organizations has greatest relevance of paper, the second largest cluster is knowledge management and so on. According to this figure it is quite relevant that the knowledge organizations is the most relevant features for most of the documents.

[h]

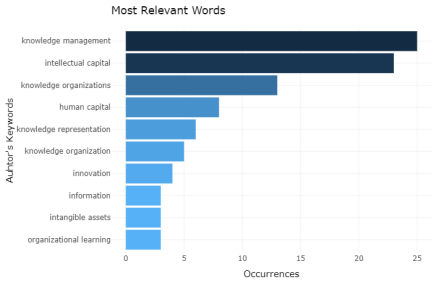


Figure 6: Most relevant words

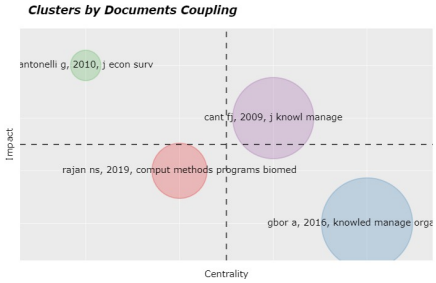


Figure 7: Clustering by documents

4 Dataset

As we have mentioned, the primary motto of the project is based on the performace od knowledge organisation, in the light of modelling human action, assessing quality and intellectual capital, we decided to analyse a data which deals with human action and assessment of its quality. Based on this motivation, we took a dataset which is known as the IBM HR Data. We obtained the dataset from kaggle. (`https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset`). In this data, there are 34 columns columns and 1470 rows. The columns contain the details of an HR and the column 'Attrition' tells us whether the employee becomes a victim of attrition or not. Thus, there are only two responses in the 'Attrition' column, viz. *yes* and *No*. Thus, this dataset provides the scope of addressing a classification problem as well as a probabilistic graphical modelling problem.

5 Models and methodology

The analysis part is divided into four parts. First we conducted an exploratory data analysis. Then We applied the probability graphical modelling, efficiency analysis and finally the problem we additionally addressed, i.e. the classification problem.

5.1 Exploratory data analysis

Explratory data analysis is the first and foremost task that we have done with this data. As this dataset is mixed up with both numerical and categorial variables, we checked which of these are categorical and which ones are numerical. Later on, we decided to go for a correlation matrix plot which could tell us the highly correlated variables. Once we get to know about the highly correlated varibles, we checked the standard features like means, medians ranges of those variables using a boxplot.

5.2 Efficiency analysis

Data envelopment analysis (DEA) is a nonparametric method in operations research and economics for the estimation of production frontiers. It is used to empirically measure productive efficiency of decision making units (DMUs). Although DEA has a strong link to production theory in economics, the tool is also used for benchmarking in operations management, where a set of measures is selected to benchmark the performance of manufacturing and service operations. In benchmarking, the efficient DMUs, as defined by DEA, may not necessarily form a "production frontier", but rather lead to a "best-practice frontier"

(a) **Constant Returns to Scale (CRS)**
The first model assumes constant returns to scale technology (CRS model). This is appropriate when all firms are operating at an optimal scale. However, note that this is quite an ambitious assumption. To operate at an optimal scale, firms should evolve in a perfectly competitive environment, which is seldom the case. The CRS model calculates an efficiency score called constant returns to scale technical efficiency (CRSTE)`https://en.wikipedia.org/wiki/Data_envelopment_analysis`.

(b) **Variable Returns to Scale (VRS)**
The second model assumes variable returns to scale technology (VRS model). This is appropriate when firms are not operating at an optimal scale. This is usually the case when firms face imperfect competition, government regulations, etc. The VRS model calculates an efficiency score called variable returns to scale technical efficiency (VRSTE) [14].
Comparison between the two models reveals the source of inefficiency. Constant returns to scale technical efficiency corresponds to the global measure of firm performance. It is composed by a 'pure' technical efficiency measure (captured by the variable returns to scale technical efficiency score) and a scale efficiency measure (SE).

(c) **Free Disposal Hull (FDH)**
The FDH model was conceptualized, formulated, developed and extended . And as such that the basic DEA models in Stata triggered off the outlier detection and the statistical inference issues in Stata, FDH model option in dea will provide Stata users with more program options to compare the results among those programs in Stata since they have their own distinct characteristics. FDH model relaxes the convexity assumption of basic DEA models. The computational technique to solve FDH program considers the mixed integer programming problem compared to the DEA model with a linear programming problem [15].

(d) **Directional Efficiency: Directional Distance Functions (DDF)**
DDF is used as the representation of production technology with multiple inputs and outputs. DDF accounts for both the reduction in the inputs or the expansion in the output simultaneously in the output direction. In DDF, Instead of a proportional contraction/expansion of the input/output vector, non-proportional resource reductions or product increasing are allowed.

5.3 probability graphical model

The promary motivation of the probability graphical model is to check and analyze the connectivity between the nodes, i.e. the features. For this model, we had to split the data into train and test set We shuffled the data at first and then we took the first 1000 rows as the train set and rest of 400 rows were considered as test set.

5.4 Logistic regression

For that classification problem, we chose to go with the most commonly used method, that is the logistic regression method. For the logistic regression problem, we divide the dataset into train and test set. 70% of the data was taken as train and the rest were taken as test. As the dataset was mixed with numerical and categorical variables, we encoded the categorical variables. Upon conduction the regression on the data, we failed to have

to satisfactory result. So we moved to an alternative option.

5.5 Neural network

As we have mentioned earlier, a poor result of logistic regression led to go for an alternative option, in this case we chose to go with a neural network model. Our model consists of two dense layer of units 128,32 and 1, respectively. For the first two dense layer, we chose *'relu'* activation function and we applied *'sigmoid'* activation function. We use binary cross entropy as loss funtion and adam optimizer. In both the cases of logistic regression and neural network, we splitted the dataset in 1:4 ratio.

6 Results and interpretation

In this section, we discuss the results we have obtained by applying aforementioned methodologies.

6.1 Exploratory data analysis

The dataset is consist of 34 variables that are categorized into two types of data:

1. **numerical variables**: Age, DailyRate, DistanceFromHome, EmployeeCount, HourlyRate, MonthlyIncome, MonthlyRate, StandardHours, TotalWorkingYears, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager.
2. **categorical variables**: BusinessTravel, Department, EducationField, Gender, JobRole, MaritalStatus, Over18, OverTime, Education, EnvironmentSatisfaction, JobInvolvement, JobLevel, JobSatisfaction, PerformanceRating, RelationshipSatisfaction, WorkLifeBalance, NumCompaniesWorked, StockOptionLevel, PercentSalaryHike, TrainingTimesLastYear.

Numerical variables are variable that has numerical values while the categorical variables consist of another type of the data(e.g. rating, boolean(true/false), 0/1, etc.) that usually categorized instead of having exact measurable values.

First we conducted a correlation plot among the numerical variables. We obtained the picture below.

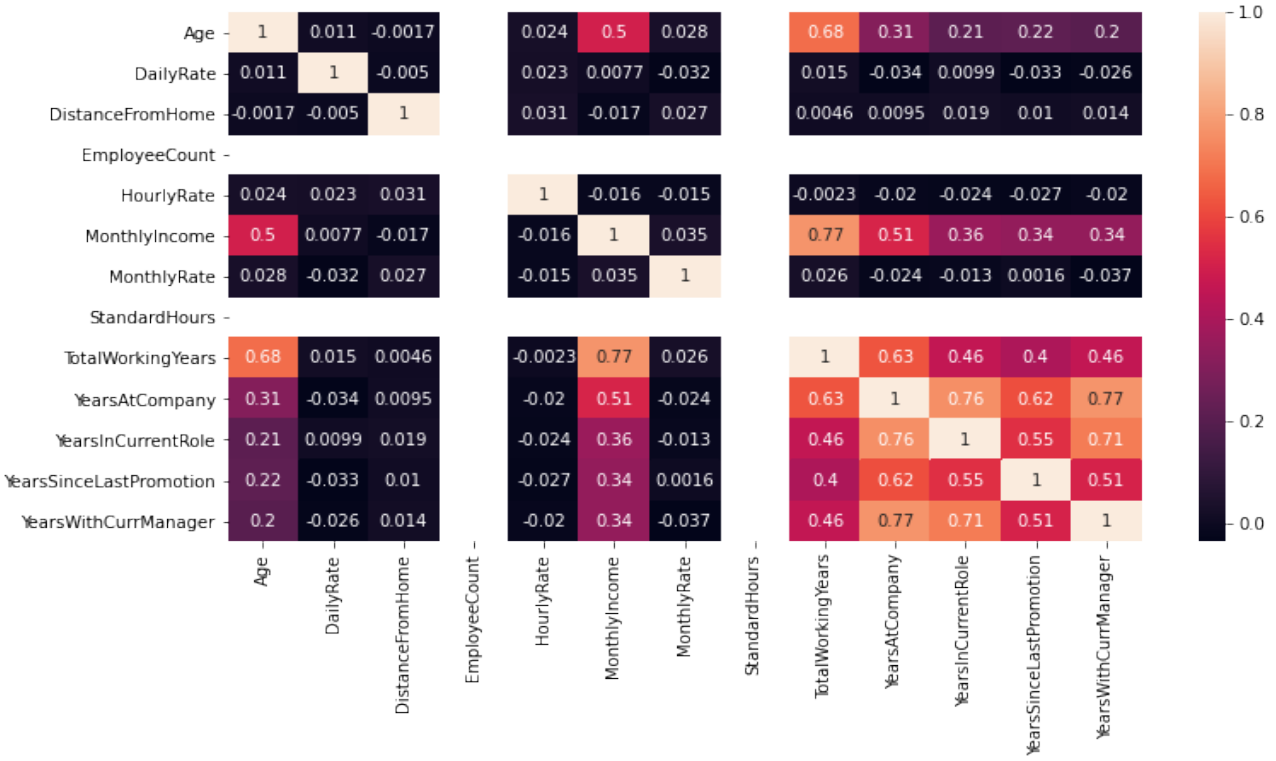


Figure 8: correlation plot among the numerical variables

Based on this correlation plots, there are some high correlation for variables: total working years and years at company. The number of total working (TotalWorkinYears) variables is highly correlated with monthly income and age (0.77 and 0.68 respectively). Moreover, the number of working at company (YearsAtCompany) is highly correlated with: the number of year working with current manager, number of years in current role, and total working years (0.77, 0.76, and 0.63 respectively).
As we have observed some highly correlated variables, we intended to plot the violin plots of the variables. The violin plots are shown below.

[h]

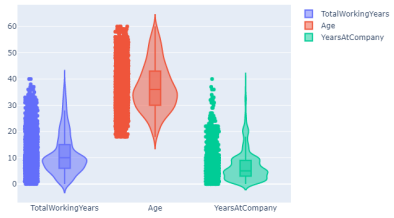


Figure 9: violin plot for numerical variables 1

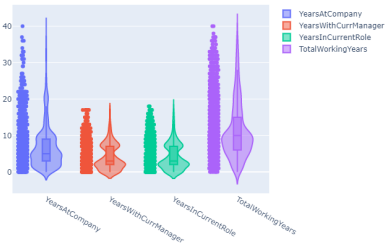


Figure 10: violin plot for numerical variables 2

In the figure 11 and 12, we represent the business travel, job satisfaction, marital status and job level in different bar plots. We see that most of the employees travel rarely and on the other hand, in each of the case of job satisfaction, marital status and job level, most of the employees' attrition is biased towards 'No'.

[h]

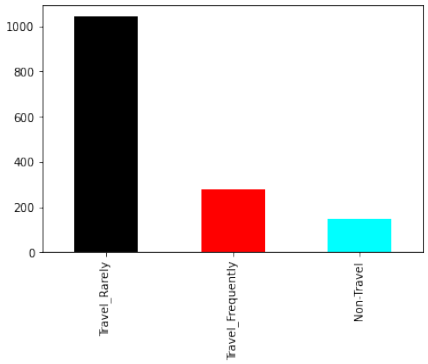


Figure 11: bar plot on business travel

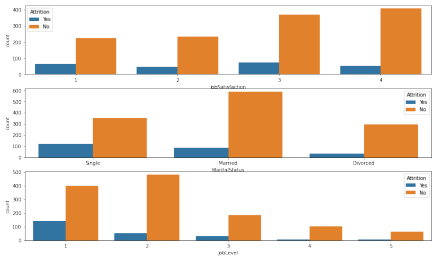


Figure 12: bar plot on job satisfaction, marital status and job level

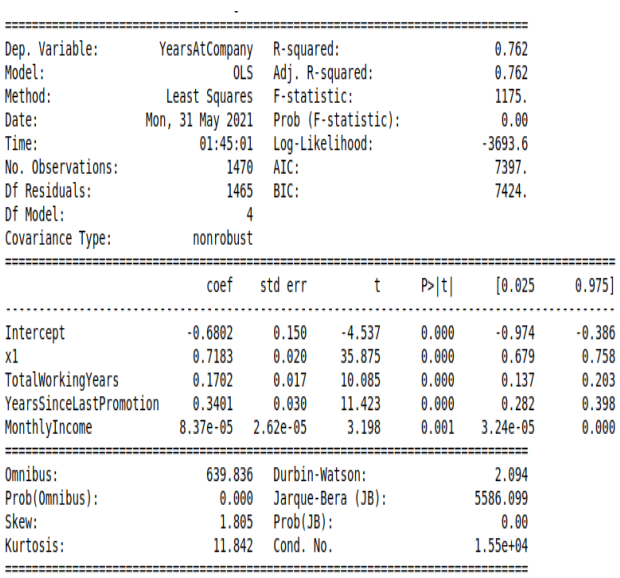
6.2 Efficiency analysis

For the efficiency analysis, We utilized some of the numerical variables from the dataset to find the relation between input and output and analyze the efficiency. By considering the correlation matrix shown in Figure 8, we conducted the inputs and output variable as shown in Table 1.

Table 1: Effeciency analysis inputs and outputs variables

Output	**Inputs**
YearsAtCompany	AvgCurrPosition, TotalWorkingYears, YearsSinceLastPromotion, MonthlyIncome

Before doing the analysis, We did some testing to make sure that our results will be reliable. We can see some statistical metrics for hypothesis testing in Figure 13



Dep. Variable:	YearsAtCompany	R-squared:	0.762
Model:	OLS	Adj. R-squared:	0.762
Method:	Least Squares	F-statistic:	1175.
Date:	Mon, 31 May 2021	Prob (F-statistic):	0.00
Time:	01:45:01	Log-Likelihood:	-3693.6
No. Observations:	1470	AIC:	7397.
Df Residuals:	1465	BIC:	7424.
Df Model:	4		
Covariance Type:	nonrobust		

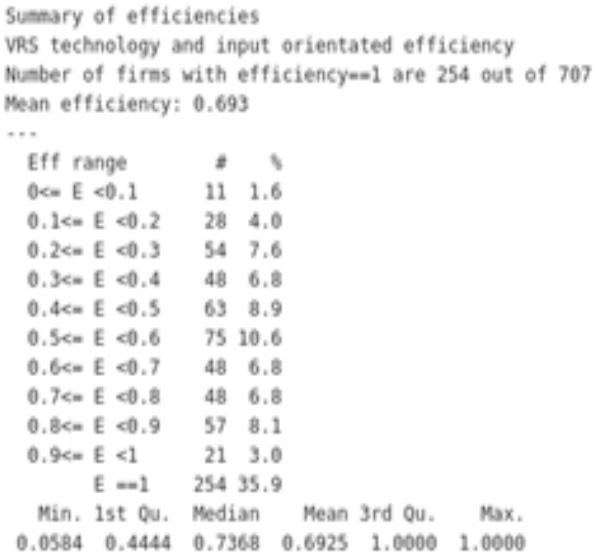
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.6802	0.150	-4.537	0.000	-0.974	-0.386
x1	0.7183	0.020	35.875	0.000	0.679	0.758
TotalWorkingYears	0.1702	0.017	10.085	0.000	0.137	0.203
YearsSinceLastPromotion	0.3401	0.030	11.423	0.000	0.282	0.398
MonthlyIncome	8.37e-05	2.62e-05	3.198	0.001	3.24e-05	0.000

Omnibus:	639.836	Durbin-Watson:	2.094
Prob(Omnibus):	0.000	Jarque-Bera (JB):	5586.099
Skew:	1.805	Prob(JB):	0.00
Kurtosis:	11.842	Cond. No.	1.55e+04

Figure 13: Hypothesis Testing

As we can see, from the result, we got the absolute t-values $|t|$ (given confidence interval with $\alpha = 0.5$)relatively high, therefore, we can inference that there are statistical significant dependence between our output and input variables.

(a) **CRS**
First, We analyze the efficiency using CRS method. The result is shown as Figure 14.



```
Summary of efficiencies
VRS technology and input orientated efficiency
Number of firms with efficiency==1 are 254 out of 707
Mean efficiency: 0.693
---
  Eff range        #    %
  0<= E <0.1      11  1.6
  0.1<= E <0.2    28  4.0
  0.2<= E <0.3    54  7.6
  0.3<= E <0.4    48  6.8
  0.4<= E <0.5    63  8.9
  0.5<= E <0.6    75 10.6
  0.6<= E <0.7    48  6.8
  0.7<= E <0.8    48  6.8
  0.8<= E <0.9    57  8.1
  0.9<= E <1      21  3.0
        E ==1    254 35.9
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0584  0.4444  0.7368  0.6925  1.0000  1.0000
```

Figure 14: CRS result

As we can see from the result above, For CRS model, 35.6% (252 employees) of the overall employees, which mean they have the most time working at the company (the loyal employees).

(b) **VRS**
Secondly, We analyze the efficiency using VRS method. The result is shown as Figure 15.

```
Summary of efficiencies
VRS technology and input orientated efficiency
Number of firms with efficiency==1 are 254 out of 707
Mean efficiency: 0.693
---
  Eff range        #    %
  0<= E <0.1      11  1.6
  0.1<= E <0.2    28  4.0
  0.2<= E <0.3    54  7.6
  0.3<= E <0.4    48  6.8
  0.4<= E <0.5    63  8.9
  0.5<= E <0.6    75 10.6
  0.6<= E <0.7    48  6.8
  0.7<= E <0.8    48  6.8
  0.8<= E <0.9    57  8.1
  0.9<= E <1      21  3.0
        E ==1    254 35.9
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0584  0.4444  0.7368  0.6925  1.0000  1.0000
```

Figure 15: VRS result

For the VRS model, 35.9% (254 employees) of the overall employees, which means they have the most time working at the company (the loyal employees). Compared to the result from CRS, this model is resulting in better performance.

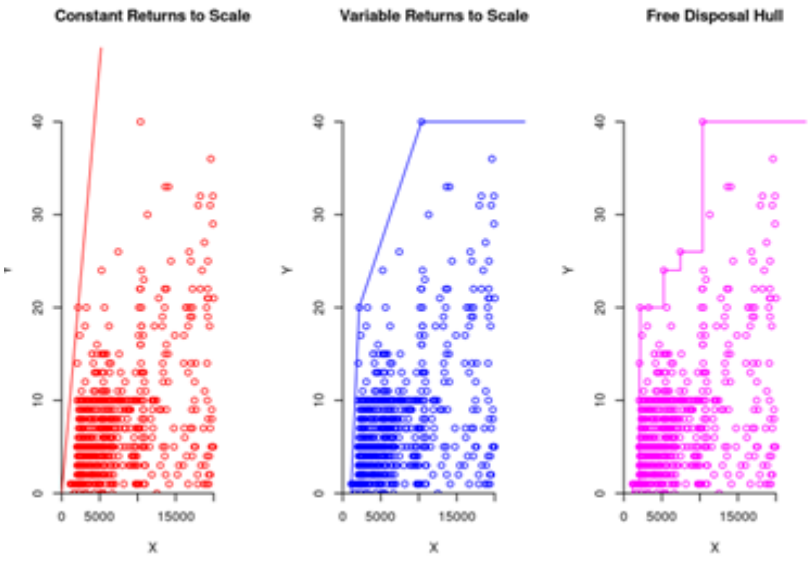


Figure 16: CRS, VRS, and FDH result

(c) **CRS, VRS, and FDH**
These plots (Figure 16) are the graphical representation of the benchmarking for data development analysis. From the result above, we can conclude that our predictors (Xs variables) as the suitable for productivity and efficiency to our output (Y variable).

(d) **DDF**
Next, We applied the directional efficiency method using Directional Distance Functions (DDF). and the results are shown in Figures 17 and 18

The first result is yielded by using the variables from this dataset whereas the later one used randomly selected scalar values to direct the model. By comparing both results, we can conclude that the directional

	X1	X2	X3	X4	peer1	peer2	peer3	sx1	sx2
1	1.0000000	1.0000000	1.0000000	1.0000000	665	NA	NA	2.50	0.000000
2	0.3333333	0.3333333	0.3333333	0.3333333	52	665	NA	0.00	0.000000
3	NaN	0.4228464	NaN	0.4228464	675	NA	NA	0.00	4.074157
4	NaN	1.0000000	NaN	1.0000000	675	NA	NA	0.00	0.000000
5	0.4117647	0.4117647	0.4117647	0.4117647	52	665	NA	0.00	0.000000
6	1.0000000	1.0000000	1.0000000	1.0000000	665	NA	NA	0.25	0.000000

	sx3	sx4	sy1
1	1.0000000	4063.5000	0
2	0.3166667	801.6267	0
3	0.0000000	0.0000	0
4	0.0000000	1564.0000	0
5	2.4888824	1248.0424	0
6	4.0000000	2377.7500	0

Figure 17: DDF directed using inputs

	e.1	e.2	e.3	e.4	e1.1	e1.2	e1.3
1	1.0000000	1.0000000	1.0000000	1.0000000	1.000000e+00	1.0000000	1.000000e+00
2	0.3333333	0.3333333	0.3333333	0.3333333	-2.220446e-16	0.6666667	-2.220446e-16
3	NaN	0.4228464	NaN	0.4228464	NaN	1.0000000	NaN
4	NaN	1.0000000	NaN	1.0000000	NaN	1.0000000	NaN
5	0.4117647	0.4117647	0.4117647	0.4117647	6.411453e-02	0.6146354	6.411453e-02
6	1.0000000	1.0000000	1.0000000	1.0000000	1.000000e+00	1.0000000	1.000000e+00

	e1.4	e2.1	e2.2	e2.3	e2.4	X1	X2
1	1.0000000	1.000000e+00	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
2	0.9994233	2.220446e-16	0.9166667	0.7500000	0.9976932	0.0000000	0.6666667
3	1.0000000	NaN	1.0000000	NaN	1.0000000	NaN	1.0000000
4	1.0000000	NaN	1.0000000	NaN	1.0000000	NaN	1.0000000
5	0.9987491	6.461892e-02	0.9037108	0.7661547	0.9949989	0.06411453	0.6146354
6	1.0000000	1.000000e+00	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000

	X3	X4
1	1.00000000	1.0000000
2	0.00000000	0.9994233
3	NaN	1.0000000
4	NaN	1.0000000
5	0.06411453	0.9987491
6	1.00000000	1.0000000

Figure 18: DDF using random scalar values

efficiency works better with the selected scalar values than selecting the variables itself as direction.

6.3 Probability graphical model

Here we plot a network map among the nodes of the dataset to see the connectivity among them. After the plotting the graph, we find the following plot.

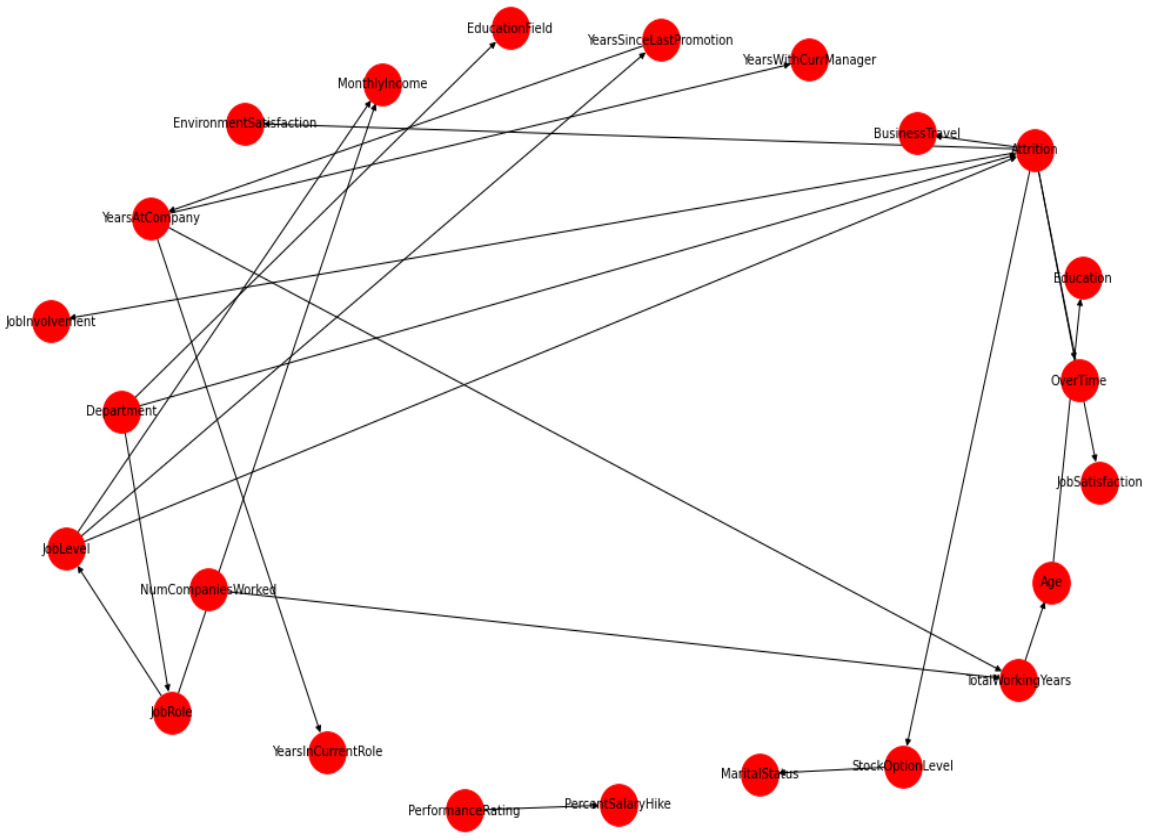


Figure 19: Probality graphical model

From the above picture, we can eshtablish a causality among the nodes. For example, we see that the node*yearsatcompany* has an influence on the total working years which in turn has an effect on age of the employee. That absolutely makes sense. But when we see that Age influences over time and overtime eventually influences the education of the employee, then that doesn't make much sense. In that case, we can force our model to follow a specific direction.

6.4 Logistic regression

We have used logistic regression in order to address the classification problem. 20 is obtained by plotting the ROC curve.

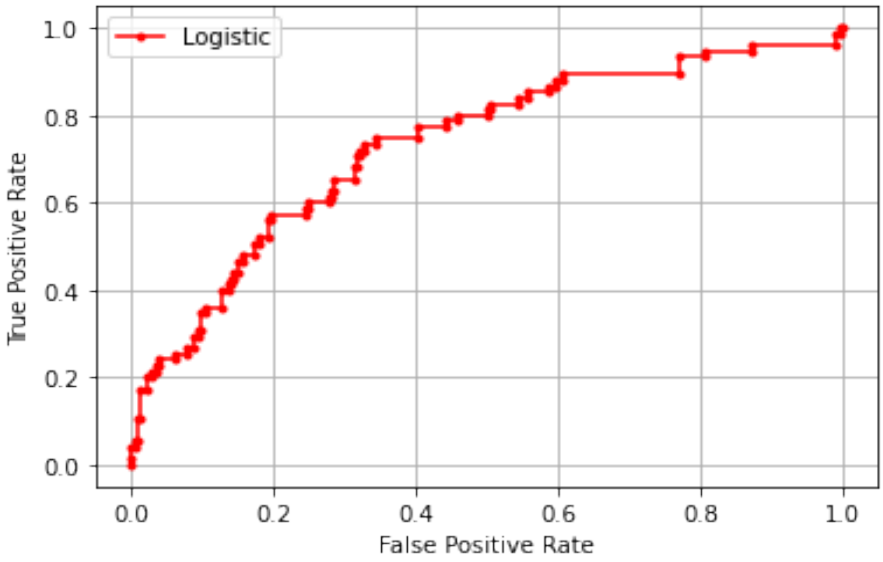


Figure 20: Logistic regression

In the case of logistic regression, we observe that the accuracy remains is 0.73 which is pretty low. So we moved to random forest.

6.5 Random forest

As mentioned earlier, having a low accuracy on logistic regression, we move to random forest technique. 21 is obtained by plotting the ROC curve.

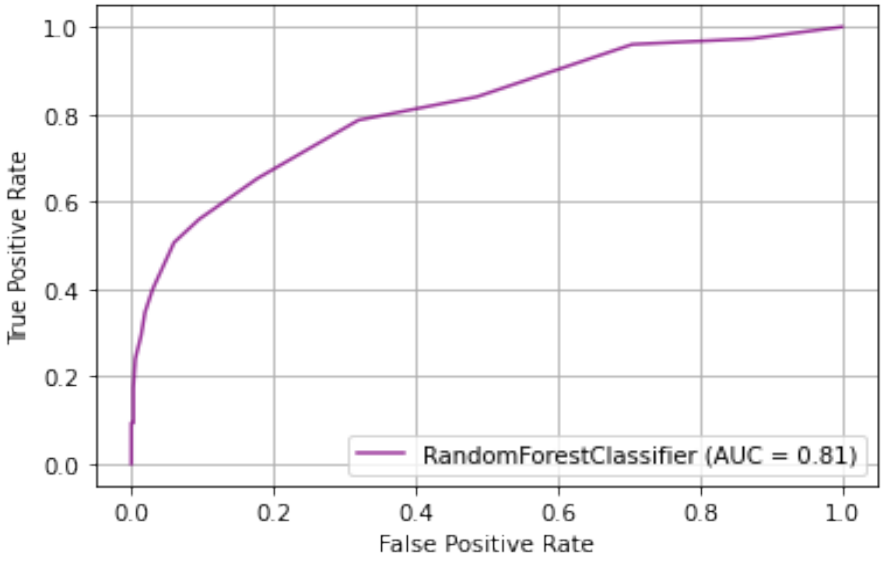


Figure 21: Logistic regression

In the case of random forest, we observe that the accuracy remains is 0.81 which is better but not so satisfactory. So we moved to random forest.

6.6 Neural netwwork

We get quite satisfactory result by using the neural network. After using the neural network with three dense layere and relu and sigmoid as activation functions, we ended up with an accuracy of around 0.89. More, the loss for both the train and validation dataset was quite low. The picture below expresses the accuracy and loss vividly.

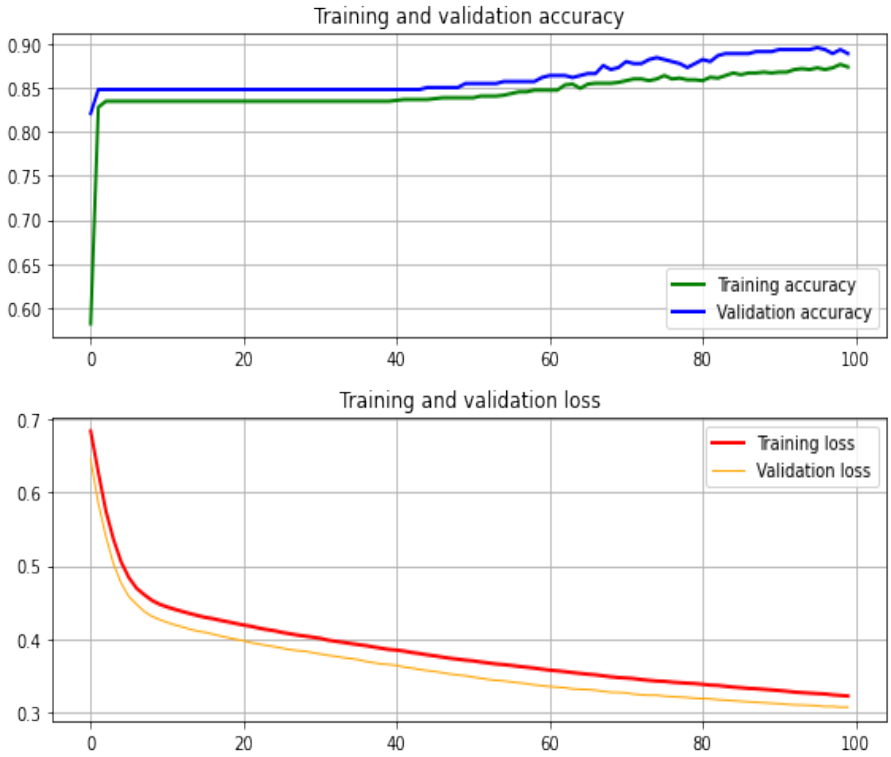


Figure 22: Loss and accuracy: Neural network

We see that at the end of the epoch, the validation loss is lower than the train loss and the valiation accuracy is higher than the accuracy of the training data.

7 Future work

It has been found that several special subjects in the investigations such as Business Management Science, Computer Science, Social Science and Decision Science covers more than 50% of the total research areas. According to VOSviewer visualization of Author's keyword using bibliographic data, in these areas there is a relationship between intellectual capital, human action with knowledge organizations, and they differ with number of citations, level of relevance which is demonstrated with multi-scaled and multi-colored circles. To strengthen the findings, the relevant words in the given areas were examined in with the help of Biblioshiny and previously mentioned areas are appeared more which proves its importance among others. Our research explored the nature of relation between the modelling human action and the knowledge organization performance from several aspects. However, accessing the information about modelling human action or observing the details and relations in this limited time interfered the effective analysis of its relation with knowledge organization. For the future investigations, these problems can be solved out, since such detailed and informative research is important in order to provide more finding about knowledge organization. We can further establish the relationship between modeling human action, assessing quality and intellectual capital on the basis of organization. Moreover, we can obtain the individual influence of the mentioned factors on the performance of knowledge organization.
So far, we have explored several aspects of the datasets. such as the relationship among the variables and their network structure, several nature and aspects of the variables and finally we managed to frame a classification challenge. Now, in this report, we have included three classification methods, viz. logistic regression, random forest and neural network. In this case, we have the scope to go deeper into the classification problem and try out other models like bayesian logistic regression, decision tree or gradient boosting. Though er have obtained best result using neural network compared to the other two, but the other unused methods can be useful too.

8 Conclusion

In terms of understanding the performance of knowledge organization, the project provides an investigation on the relationship between human resource management and knowledge organization. Obtained results by using the VOSviewer, Biblioshiny and Scopus, it is demonstrated that there is a link between intellectual capital, human activity and knowledge organization, and quality assessment is not in the spotlight as it is a factor in all of the above. Moreover, we highlight the factor of concern of our project with ontological modelling. While dealing with the data set, we observed that the classification task can be done by using different methods with a steady increase of accuracy with logistic regression, random forest and finally neural network. We have seen another important aspect as well. While modeling the probabilistic graphical modeling, we found that there are connections between few nodes which are not supposed to make much sense. This problem can be challenged by reducing the number of columns, that is, there might be few columns in the dataset which are not so significant to address the problem of classification and probabilistic graphical modeling. In a nutshell, by conducting the systematic review and the modeling on human action, we observed the imminent role of knowledge organization and human action in a vivid manner.

References

[1] G. Reinmann, *Knowledge Organization*. Boston, MA: Springer US, 2012, pp. 1686–1689. [Online]. Available: https://doi.org/10.1007/978-1-4419-1428-6_445

[2] B. Hjørland, "What is knowledge organization (ko)?" *KO Knowledge Organization*, vol. 35, no. 2-3, pp. 86–101, 2008.

[3] L. Von Mises, "Human action: A treatise on economics," 1949.

[4] P. Kamaja, K. Löytty, M. Ruohonen, and T. Ingalsuo, "Evaluation of knowledge work productivity–case distributed software development," in *European Conference on Intangibles and Intellectual Capital*. Academic Conferences International Limited, 2017, p. 142.

[5] C. BIANCALANA, F. GASPARETTI, and A. MICARELLI, "Knowledge retrieval and personalization in virtual enterprises," *management*, vol. 10, no. 27, p. 24.

[6] G. Peng *et al.*, "The estimation of enterprise's human capital operation," in *2006 International Conference on Management Science and Engineering*. IEEE, 2006, pp. 1350–1354.

[7] M. Smits and A. de Moor, "Measuring knowledge management effectiveness in communities of practice," in *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the.* IEEE, 2004, pp. 9–pp.

[8] W. E. Carnes, "Knowledge management: A state of the practice summary," in *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*. IEEE, 2002, pp. 6–6.

[9] J. Liebowitz, "Knowledge assets and their scheduling of use within organizations," in *Proceedings Intelligent Information Systems. IIS'97*. IEEE, 1997, pp. 262–265.

[10] W. Fan and N. Bouguila, "Variational learning of finite beta-liouville mixture models using component splitting," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2013, pp. 1–8.

[11] G. Hodge, D. Soergel, and M. Zeng, "Building a more meaningful web: From traditional knowledge organization systems to new semantic tools," *D-Lib Magazine*, vol. 9, no. 7/8, pp. 1082–9873, 2003.

[12] N. J. Van Eck and L. Waltman, "Visualizing bibliometric networks," in *Measuring scholarly impact*. Springer, 2014, pp. 285–320.

[13] G. Ganino, D. Lembo, M. Mecella, and F. Scafoglieri, "Ontology population for open-source intelligence: A gate-based solution," *Software: Practice and Experience*, vol. 48, no. 12, pp. 2302–2330, 2018.

[14] J.-M. Huguenin, "Data envelopment analysis (dea)," *A pedagogical guide for decision makers in the public sector, Swiss Graduate School of Public Administration, Lausanne*, 2012.

[15] R. M. Thrall, "What is the economic meaning of fdh?" *Journal of productivity analysis*, vol. 11, no. 3, pp. 243–250, 1999.