Report on

ARTIFICIAL NEURAL NETWORKS FOR SUPPLY CHAIN MANAGEMENT

Submitted By :
CHIRAG VERMA
2K19/BMBA/04

DEVANSH AGARWAAL
2K19/BMBA/06

Under the Guidance of :
DR. KAMAL GULATI
Associate Professor (Amity University)

UNIVERSITY SCHOOL OF MANAGEMENT & ENTREPRENEURSHIP

Delhi Technological University

MAY 2021

CERTIFICATE

This is to certify that Chirag Verma (2K19/BMBA/04) and Devansh Agarwaal (2K19/BMBA/06) are bona fide students of University School of Management & Entrepreneurship, Delhi Technological University and have successfully completed the project work as prescribed by Delhi Technological University in the partial fulfillment of the requirement of Masters of Business Administration (MBA), Business Analytics Program for the academic year 2019-2021.

The project work is titled as "**ARTIFICIAL NEURAL NETWORKS FOR SUPPLY CHAIN MANAGEMENT**".

Project Guide

Dr. Kamal Gulati

Associate Professor (Amity University)

DECLARATION

We, Chirag Verma and Devansh Agarwaal, hereby declare the presented report titled **"ARTIFICIAL NEURAL NETWORKS FOR SUPPLY CHAIN MANAGEMENT"** is our original work and is prepared only after the completion of our work at the organization.

We confirm that we have not used work previously produced by another student or any other person to hand in as our own.

We also confirm that this report has been undertaken for the sole purpose of partial fulfillment of Masters of Business Administration (MBA), Business Analytics Program at University School of Management & Entrepreneurship, Delhi Technological University.

CHIRAG VERMA
2K19/BMBA/04
MBA Business Analytics
University School of Management & Entrepreneurship
Delhi Technological University

DEVANSH AGARWAAL
2K19/BMBA/06
MBA Business Analytics
University School of Management & Entrepreneurship
Delhi Technological University

ACKNOWLEDGEMENT

A successful project is the result of teamwork and coordination that includes not only the group of developers who put forth the ideas, logic and efforts but also those who guide them. We will be using this opportunity to express our gratitude to everyone who supported us throughout the course of this report.

We are thankful for all the knowledge, guidance and support imparted by Dr. Kamal Gulati at any time we needed.

We highly appreciate the faculty members of University School of Management & Entrepreneurship, Delhi Technological University for their guidance and constant supervision as well as for providing necessary information throughout the course of this project.

We would also like to express our gratitude towards our parents and friends for their kind cooperation and encouragement which helped us in the completion of this project.

CHIRAG VERMA

DEVANSH AGARWAAL

ABSTRACT

In simple words a supply chain is getting the raw materials, converting them into a finished product and delivering that to the end customer. The product can be tangible (a good) or intangible (a service) or even a combination of both. Basically it is the process of manufacturing and delivering. Supply chain management is simply handling the supply chain in order to have optimum outputs. Artificial Intelligence (AI) is the most modern form of machine learning and is what is considered as a hot topic of research in today's world. Simply put, AI is mimicking the human intelligence and reactions that would happen in a given situation. Artificial Intelligence is a booming field and so is Supply Chain Management, and in recent years researchers and companies such as consultancy firms have been trying to integrate both together in order to have optimum supply chains either for themselves or for their client(s).

In this report, we'll be having an introduction to the fields of supply chain and artificial intelligence coming together, learning about previous studies where this combined concept has been worked upon, talking about various machine learning techniques for model creation, working on a real world supply chain shipment data set using the deep learning tool of Artificial Neural Networks (ANN) using the Python programming language for creating a prediction model and interpreting the results derived such as model accuracy.

TABLE OF CONTENTS

CERTIFICATE i

DECLARATION ii

ACKNOWLEDGEMENT iii

ABSTRACT iv

TABLE OF CONTENTS v

INTRODUCTION 1

1.1 Industry Profile 1

1.2 Organization Profile 6

1.3 Objective of the study 11

LITERATURE REVIEW 12

RESEARCH METHODOLOGY 18

RESULTS 44

FINDINGS AND RECOMMENDATIONS 48

LIMITATIONS OF THE STUDY 50

BIBLIOGRAPHY 51

ANNEXURE 52

PLAGIARISM REPORT 54

1. INTRODUCTION

1.1 Industry Profile

Supply Chain refers to the process of obtaining the raw materials from the source which will be transported to the place of manufacturing, passing through each and every step requiring different set of people with different sets of skills and being converted into the finished goods and then finally transported to the destination point, i.e., the sellers shop or even the end customer. In simple words it is the process of manufacturing and selling goods. In today's world an optimum supply chain is the one that every industry desires. This optimum supply chain helps the firms in having a faster production cycle at the least possible cost.

Supply Chain Management (SCM) as the term suggests is the management of the flow of goods and services such as movement and storage of raw materials, work-in-progress inventory and finished goods from point of origin to the point of consumption. It is a broad term that consists of various activities such as sourcing, production, product development, operations and logistics. In today's business world companies are focusing more and more in the field of supply chain management in order to reduce risks.

Table 1.1 Flows in supply chain

Information Flow - ● Invoices ● Sales Literature ● Specifications ● Receipts ● Orders ● Rules & Regulations	Primary Product Flow - ● Materials ● Components ● Supplies ● Services ● Finished Products
Primary Cash Flow - ● Payment of products ● Supplies	Reverse Product Flow - ● Returns for repair ● Replacements ● Recycling ● Disposals

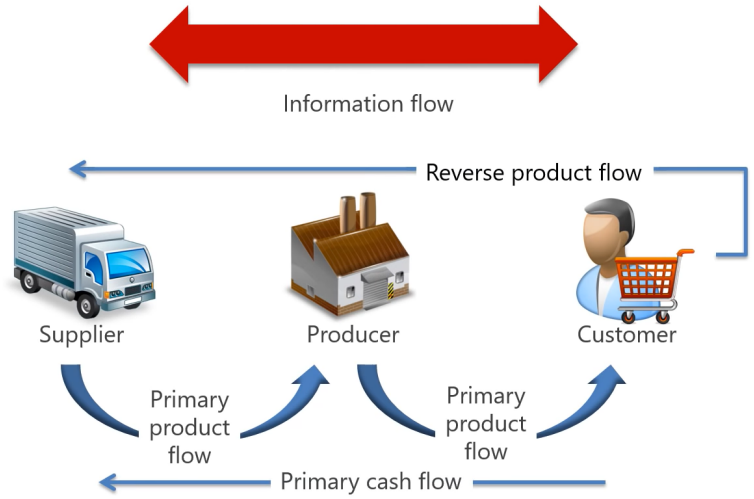

Fig. 1.1 Basic Supply Chain

A basic supply chain consists of three components, i.e., the supplier who provides goods and services or a person or organization with whom the buyer does business as they provides materials, energy, services or components for a product or service; second the producer, who receives the components from the supplier to produce the finished goods or service; third the customer, the one who receives the finished product.

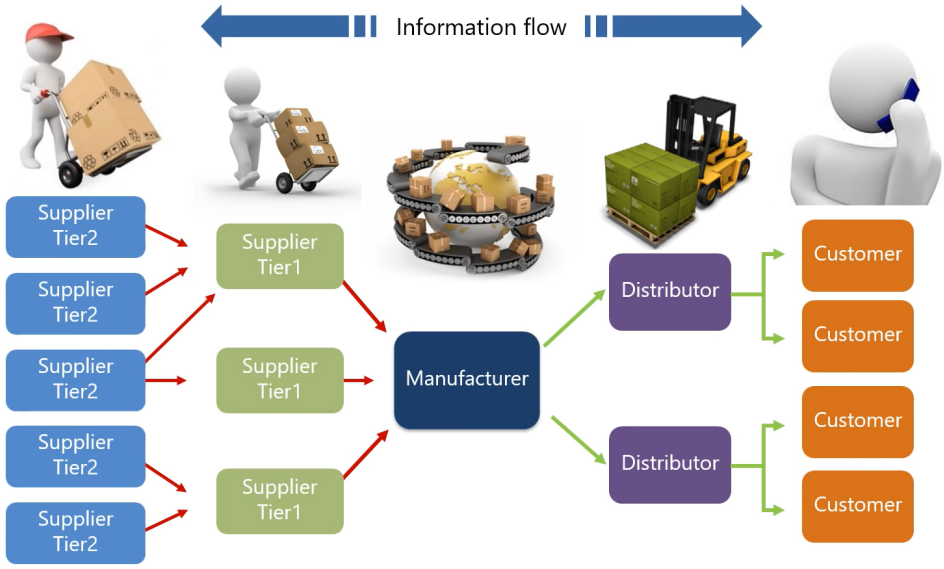

Fig. 1.2 Supply Chain in Manufacturing

In case of manufacturing there can be different levels of suppliers as a supplier can have various suppliers who will provide them with the goods. This chain can go on till even Tier 4 suppliers or even more. The same can goes with distributors as there can be various channels of distribution of the finished goods till they reach the final consumer.

Fig. 1.3 Supply Chain in Services

Supply chain is not only followed in the manufacturing sector but also in the service sector such as electricity, software, legal advisor, federal government, etc. The companies in the service sector receive their products, services and supplies from other firms and then provide their services to home customers, commercial customers and other utilities.

In today's world supply chain management plays a major role in the firms' success. Top global brands such as Amazon, Coca-Cola, Zara, Unilever, etc. are examples of efficient supply chain management that helped them in optimizing profits as well as meeting customer satisfaction as they welcomed the opportunity for growth and future investment by opting to supply chain management.

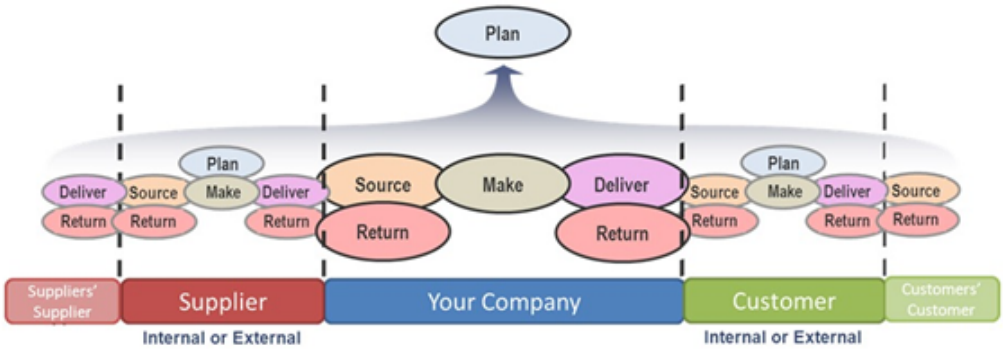

Fig. 1.4 SCOR Model

Supply chain management has a few models and tools that are needed by the firms in today's world in order to have an effective supply chain. The most commonly used model of supply chain management is known as SCOR Model. It refers to the Supply Chain Operations Reference Model that tells about how organizations use supply chains which can span from supplier's supplier to customer's customer and includes five stages namely, plan, source, make, deliver and return.

Supply chain and logistics follows a sales execution strategy known as Route to Market (RtM) strategy which provides a roadmap of transportation of goods and services from factory to the end users in an effective and efficient manner. There are several route-to-market methods that can be followed such as selling online, direct sales, selling to retailers, selling to wholesalers, remote sales, etc.

A set of predefined rules by the International Chamber of Commerce (ICC) known as International Commercial Terms or simply Incoterms are followed by the governments and legal authorities worldwide are intended to clearly communicate the tasks, costs, and risks associated with the global or international transportation and delivery of goods.

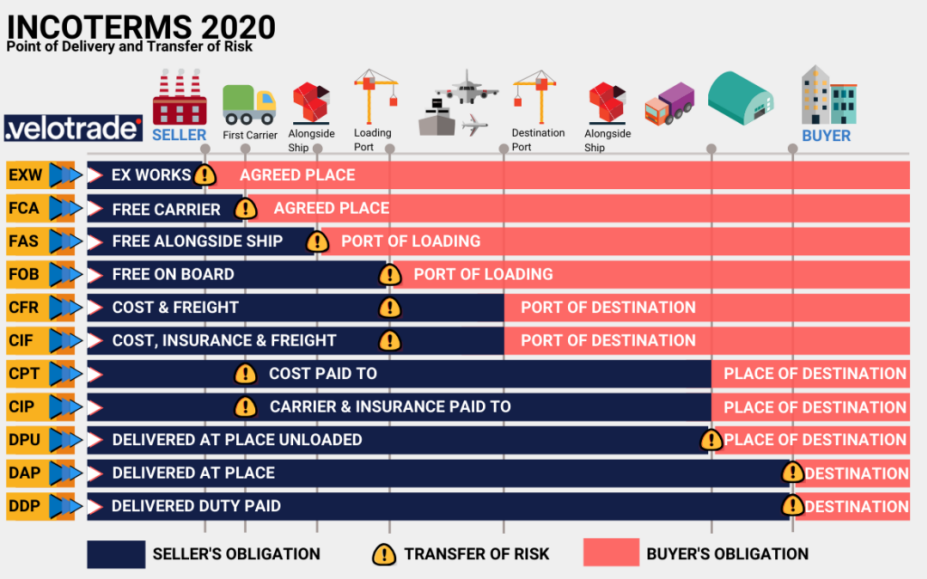

Fig. 1.5 Incoterms 2020

There is a total of 11 Incoterms rules –

RULES FOR ANY MODE OR MODES OF TRANSPORT (7)

- EXW - Known as Ex Works
- FCA - Known as Free Carrier
- CPT - Known as Carriage Paid To...
- CIP - Known as Carriage and Insurance Paid to...
- DAT - Known as Delivered At Terminal
- DAP - Known as Delivered At Place
- DDP - Known as Delivered Duty Paid

RULES FOR SEA AND INLAND WATERWAY TRANSPORT (4)

- FAS - Known as Free Alongside Ship
- FOB - Known as Free On Board
- CFR - Known as Cost and Freight
- CIF - Known as Cost, Insurance and Freight

1.2 Organization Profile

One of the top firms in the world in the domain of logistics and supply chain management is Enchange®.

Founded in 1993 by Michael Thompson, Enchange Limited is an international supply chain and Route to Market (RtM) consultancy firm registered in England and Wales with its headquarters located in London. The company has experience in various industries such as -

- FMCG (Fast Moving Consumer Goods)
- Brewing & beverages
- Pharmaceuticals
- Retail
- Automotive, Industrial & Engineering
- Private Equity & Financial
- Telecommunications

The company has a clientele spread over 70 countries with 100 different clients for which they've developed more than 600 supply chain related projects. The company has worked with small players for region specific markets as well as with large multinational organizations. Some of the major organizations with whom Enchange has worked are -

- British American Tobacco
- Unilever
- United Biscuits
- Lever Brothers
- Coca-Cola

- Halwani Brothers
- Manifatture Sigaro Toscano
- Heineken
- SABMiller
- Guinness
- Brau Union Romania
- Arysta LifeScience Corporation
- The Candel Company Limited
- GlaxoSmithKline
- Novartis International AG
- Polpharma SA Pharmaceutical Works
- AKRIKHIN
- Sicomed
- Labormed Alvogen
- Terapia Ranbaxy
- BIOFARMA Pharmaceutical
- Munir Sahin Ilac Sanayi Ve Ticaret AS
- Betasan
- Orange Business Services
- Celtel
- Connex-Vodafone
- Ericsson
- Tata Motors
- Royal Dutch Shell
- Firestone Tire and Rubber Company
- Romstal
- Valrom Industrie
- European Bank for Reconstruction and Development
- Department for International Development

- Global Finance
- GED Capital
- Romanian - American Foundation (RAF)
- Advent International

The firm provides four major services to its clients in order to improve their Route to Market (RtM) and Supply Chain strategy. These are -

a. Route to Market & Distribution -

A 20 steps Route to Market program divided into four phases combining decades of commercial insights and operational execution that enables the firms to have their current RtM strategies analysed and implement solutions in order to improve effectiveness and efficiency resulting in increasing profits.

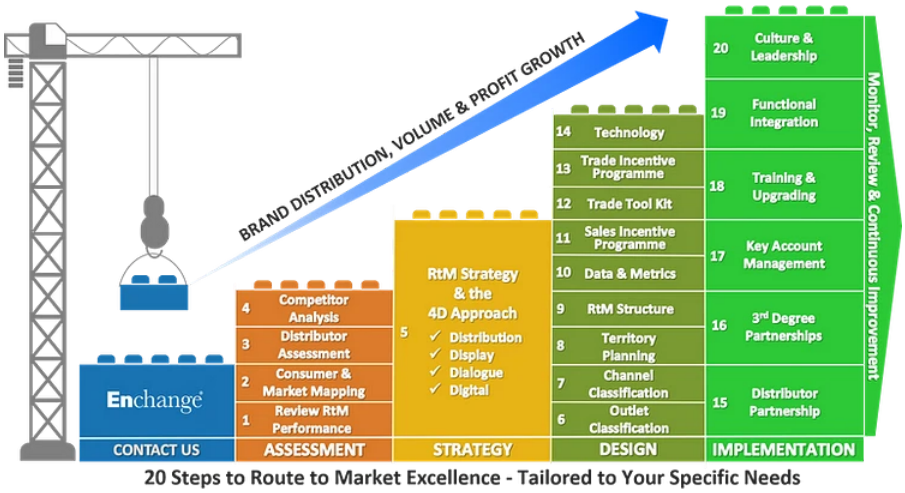

Fig. 1.6 Route to Market Excellence

b. Supply Chain Analytics -

Using SupplyVue, a supply chain analytics solution made in collaboration with Concentra that uses the data provided in order to analyse our supply chain and

provides different analytical tools and dashboards that can be modified as per the scenario that helps in efficient decision making.

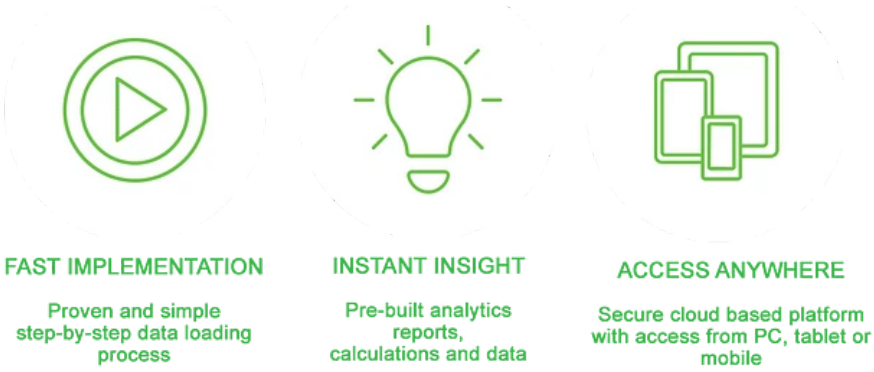

Fig. 1.7 Usage of SupplyVue

c. Supply Chain Transformation -

The Enchange Supply Chain House model is based on the SCOR model developed by experts having knowledge of local, regional and international markets and provides us with clearly defined supply chain and understanding its potential.

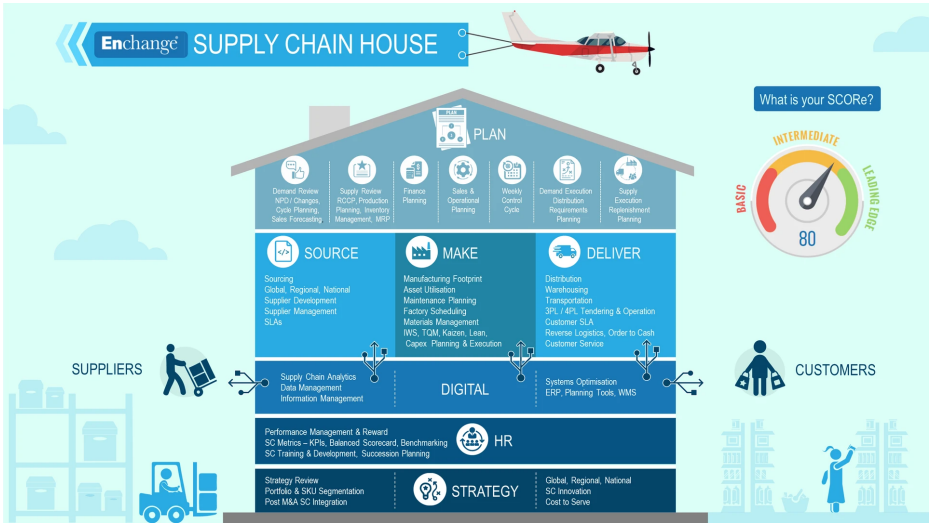

Fig. 1.8 The Enchange Supply Chain House

d. Supply Chain Talent Hub -

Operational and Supply Chain experts with decades of experience in consultancy in various industries ready to engage in activities with the clients as they support key projects, which will help in initiating managerial and impactful changes within the organization.

Fig. 1.9 Homepage of the firm's website

1.3 Objective of the study

The primary objective of this report was to complete the project. As students of an MBA program every person has to complete a project dissertation in their final semester where they have to plan and execute a project by investigating a decided topic and making a report on their findings. Now we have completed our project on use of Artificial Intelligence in Supply Chain Management.

There are various other objectives, such as -

- To know the concept of supply chain.
- To understand the working of supply chain management.
- To understand the concept of Artificial Intelligence (AI).
- To study various techniques Machine Learning Techniques.
- To know how to use AI in optimizing supply chain management.

2. LITERATURE REVIEW

In the era of globalization it has been observed that people have more freedom in choosing the best alternative products. This is due to the difference in the product features, designs, etc. One of the major factors that affect the supply chain of a firm is its product design as they majorly play a role in determining the product demand. In reality it also affects the manufacturing process, quality, quantity, transportation, time and most importantly production cost. As the organizations try to achieve flexibility in product manufacturing they are continuously decreasing the processes to be carried out by them and rather outsource activities to other firms that have capability of producing the goods effectively and efficiently.

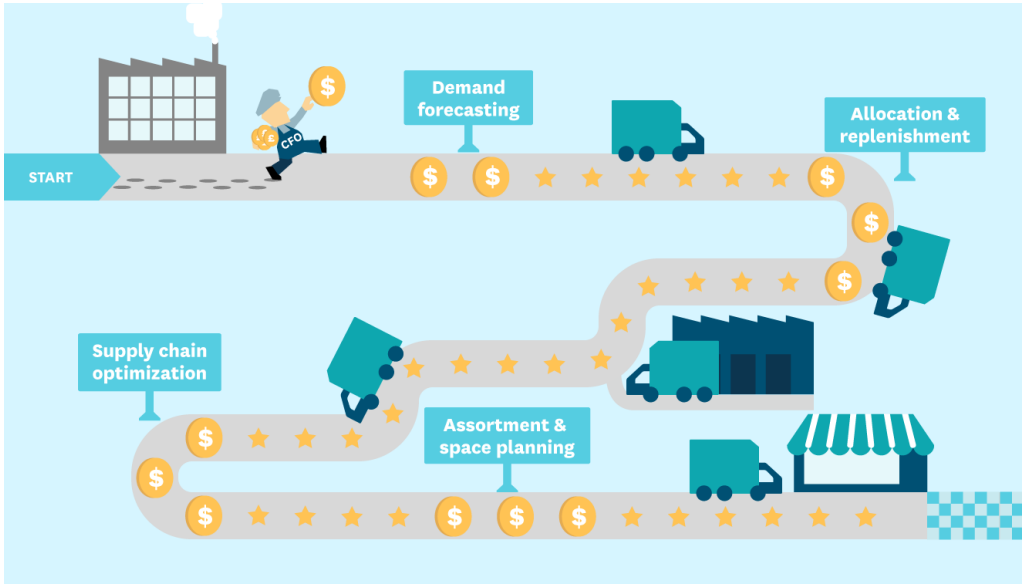

Fig. 2.1 Supply Chain Transformation

There are many firms that have determined that their sole purpose is to manage supply chain for other firms, manufacturing or service, both. They look after allocating the various activities such as who will be the supplier of raw materials, what stages will be needed in order to create the good, how it will be transferred from one stage to another or transported from one place to another, how the main supplier will assemble the product or service, how the ownership will be transferred to the buyers' hands, etc.

These firms usually use different supply chain models which are being run by artificial intelligence in order to find if there are any discrepancies or not and try to optimize the supply chain.

There have been a few studies earlier that have studied the use of artificial intelligence in order to optimise supply chain management.

As per the study, **“Artificial intelligence in supply chain management: theory and applications”** the author has initially talked about the concepts of supply chain and supply chain management and how it is becoming more information intensive. The author then talks about artificial intelligence, what it is, how it mimics human understanding and decision making processes and how artificial intelligence can help in optimizing supply chain management by predicting the future outcomes by studying past data. The author then defines the problem scope in three categories regarding decision making namely; first, strategic decision that are usually executive-level decisions made having a long term view in mind regarding strategic partnerships, investments, etc., second, tactical decisions that are usually made by the mid-managerial level people that are going to be in effect for a considerable time period and focuses on matters like demand and supply, inventory planning, etc, and operational decisions that are made by the operational managers for the short term day to day activities relating order delivering, etc. The author has then identified and explained various categories of artificial intelligence such as artificial neural network, rough set theory, machine learning (and its subcategories like concept learning, decision tree learning, perceptron learning, bayesian learning and reinforcement learning), expert systems (its components such as knowledge base, inference engine, justifier and user interface), genetic algorithms, fuzzy logic (its components such as linguistic variables, linguistic values, fuzzy sets, membership functions and fuzzy IF-THEN rules) and agent-based systems. The author then explains how these artificial intelligence techniques have been implemented in supply chain settings such as

inventory control & planning, transportation network design, purchasing & supply management, demand planning & forecasting, order-picking problems, customer relationship management and e-synchronised supply chain management. The author has then analysed and compared various other studies where different artificial intelligence techniques have been used on real-world datasets and their findings and concluded that agent-based system is the most popular artificial intelligence tool in tackling various supply chain problems as it is most linked tools with different supply chain areas.

The study done in **"Artificial intelligence in supply chain management: A systematic literature review"** the authors have initially explained the concepts of artificial intelligence, how it is being helpful in improving human life, how companies are now shifting towards a more AI monitored sophisticated system in order to improve their functioning, importance of artificial intelligence in the industry, and how does AI contribute in supply chain management studies by focussing on production, marketing, logistics and supply chain management itself. The authors then did their research in a systematic manner by opting to a five-phase research methodology. In the first phase the authors did a pilot search by identifying five major journals, namely, ScienceDirect, Emerald insight, JSTOR, Wiley online library and Taylor & Francis where they studied the database of these journals in order to have better understanding of the topic on which they were researching and then forming the research question and sub-research questions. In the second phase, i.e., locating the studies, the authors identified the relevant publications from the five mentioned journals' databases by using a specific set of keywords for their string searches as a result of which the authors identified 758 potential articles. In the third phase, study selection and evaluation, the authors in order to eliminate irrelevant articles, read the abstracts of the articles in order to find whether they're working on the same grounds or not which resulted in reducing the number to 64 articles. In the fourth phase, the authors broke the selected 64 articles on the basis of the specific set of characteristics that would answer the research question which

resulted in the breakup of the articles where 14 were related to marketing, 23 related to production, 6 related to logistics and 21 related to supply chain. These papers were further divided on the basis of the artificial intelligence techniques used, such as artificial neural networks, agent-based systems, support vector machines, association rule mining, k-means clustering, etc. on the basis of which it was found out that ANN is the most popular AI technique used. In the fifth phase, the authors have then as per their study and analysis of the various articles have answered the dub-research questions that were formed in the beginning to support the main research question and concluded that the artificial neural networks is the most preferred technique used in supply chain management with fuzzy logic coming after it.

In the study, **"Forecasting Model of Supply Chain Management Based on Neural Network"** the author has initially explained the concept of neural networks, what the structure of a neural network comprises of, how it is based on two concepts feedforward and backpropagation, characteristics of neural network and how a neural network works by explaining its process, and finally putting forward a few limitations of neural network. The author then explains the concept of supply chain management and how neural networks help supply chain management in three main areas, namely, optimization, forecasting and decision support. The author then takes the case of a bicycle market of a certain region and then applies two models on the data set, Backpropagation Neural Network Model and Linear Regression Model for predicting the market demand of the bicycles and comparing them with the actual demand finding out that neural network model has higher accuracy of the two models. The author concludes that the neural network is a preferred model for predicting and optimizing supply chain management as it is able to adapt to the deficiencies and keep on improving from itself.

According to the study, **"A Study Regarding the Possibility of Optimizing the Supply Batch using Artificial Neural Networks"** the author talks about optimization of inventory management systems using artificial neural networks. She initially states the

two problems associated with inventory management, namely, issuing of new purchases and optimal batch sizes. The author then talks about various papers that have used artificial intelligence tools such as fuzzy sets, neural networks, etc. to solve these two issues. It is stated that a combination of both statistical methods and machine learning is also used in order to resolve the two stated issues. The author then explains that statistical methods need historical data for prediction and if they are not present due to situations like a new company, new product inventory, social and economic environmental changes, etc. then the statistical methods are invalid. The author then provides a brief description of the statistical model explaining the formula and working of the statistical models for inventory management that will determine the optimum batch sizes. The author then explains the concept of neural models for determining batch size and gives the real world example by using a three layer perceptron model and using the training data. The author then concludes that the neural network model is the best used in such scenarios by having a comparative study.

The authors in the study, **"Improving Supply Chain Visibility With Artificial Neural Networks"** have firstly addressed the challenges regarding the vulnerability of the supply chain in predicting the capacity of simulated supply chains by anticipating when and by how much will be the requirement for the order of supply. The authors then explain the concepts of supply chain, risks associated with it and how some organizations are able to cope with the risks. The authors then explain the concepts of artificial intelligence and how it profits organizations by using techniques such as artificial neural networks by mimicking human behaviour. In order to know how artificial neural networks will be able to help in predicting the capacity of supply chain, the authors use a multi-echelon supply chain simulation model using the Simul8® software where a following set of parameters is being set such as supplier, buyer, production and distribution, and afterwards the models are feeded to the artificial neural networks which is developed using the MATLAB software. In the artificial neural network 3 layer perceptron is used where two different experiments are done. In experiment 1, the

authors try to anticipate the quantity of the orders to be sent when order is made and expected time to fulfill the orders, which came out to be 99-98% accurate. In experiment 2, the authors see which entity will reach its re-order point and place orders which came out to be as low as 75% accurate. The authors conclude saying that artificial neural networks can be used to predict any disruptions in the supply chain.

3. RESEARCH METHODOLOGY

Artificial intelligence (AI) is the development and implementation of a computer system that is able to make decisions on the basis of its own logic and reasoning. The machine is able to make the decisions after it does the visual, vocal and sentiment analysis of the observed data.

On a daily basis, people have been interacting with AI systems. E-commerce companies like Amazon, large social media companies like Facebook and world renowned search engine Google have been using the AI tools for quite some time now. On Amazon, when a person have searched for a few products that can be linked together and looking just under them, it says, “Here are some products recommended for you”, it’s not just a coincidence that the recommended item is similar to the searched item or works together with it, it is because of AI that people see those recommended products. Popular digital personal assistants like Alexa (Amazon), Bixby (Samsung) and Siri (Apple), are all using artificial intelligence in order to ease everyone's living.

Artificial Intelligence consists of various machine learning algorithms such as -

1. Artificial Neural Networks (ANN)

   An artificial neural network (ANN) is a machine learning technique that analyses the data set provided to it by mimicking the function of the and decision making process of a human brain and then finds the underlying relationships among the data points. An ANN consists of one or more layers of nodes which are connected to each other with dendrite like structures for sending inputs. These nodes are also known as perceptrons that use supervised learning in order to have an output. The perceptron uses multiple linear regression and an activation function.

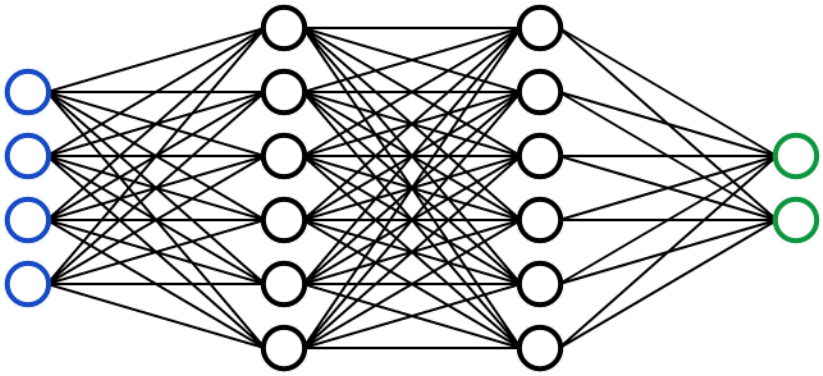

Fig. 3.1 Structure of Artificial Neural network

Due to its self-learning capabilities ANN is able to analyse more and more data provided to it and which inturn helps in having better results. ANNs use a set of learning rules called forward pass, an abbreviation for forward propagation, and backpropagation, an abbreviation for backward propagation of error, to perfect their output results.

2. Rough Set Theory -

Rough Set Theory is a data mining technique where the information system (IS) is analysed using the concept of approximations where each row in the data table provided is being analysed in order to extrapolate patterns, feature extractions, etc. It consists of various concepts such as indiscernibility where the tuples having the similar values are clubbed together, setting lower approximation meaning the indiscernibility positively belongs to the target set and upper approximation results meaning the indiscernibility possibly belongs to the target set and then combining the two results in deciding which all indiscernibility belong to the target set.

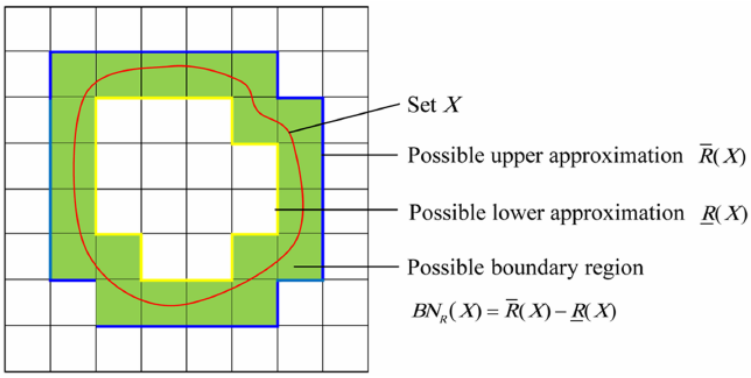

Fig. 3.2 Rough Set Approximation

3. K-Nearest Neighbour

K-Nearest Neighbor algorithm (K-NN) is a non-parametric method for the purpose of pattern recognition by using classification and regression. K-NN algorithm compares the previously recorded instances in its database and compares them with the new data provided for classification and it keeps on learning. The classification process in K-NN is done by having a majority of the values/neighbours together in a cluster. If K = 1, then the class is the single nearest neighbor. Euclidean distance which is the shortest distance between two points in a straight line is used for assigning weights to a neighbour by multiplying the value with 1/d. The closely located, i.e., the nearest values after determining the clusters helps in pattern recognition.

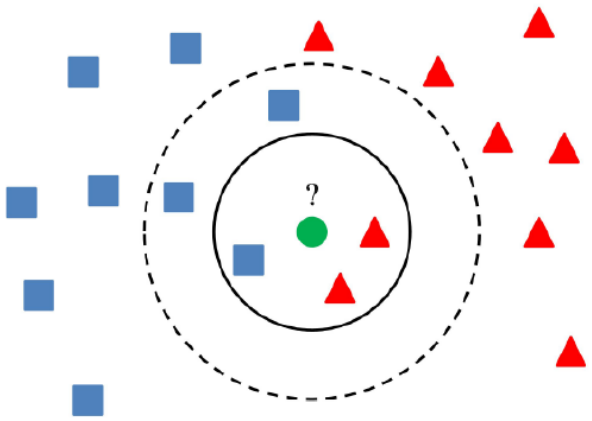

Fig. 3.3 K-Nearest Neighbour

4. Naïve Bayes Classification -

Bayes' Theorem which calculates the probability of an event happening or not based on the knowledge provided priorly has a further extension where the algorithm is based on strong assumptions whether one thing will happen or not. The decisions in the Naïve Bayes classification are made on the basis of conditional probability. For the purpose of creating the classification model the data set is divided into two subsets, namely, feature matrix and response vector. The feature matrix consists of the values of the independent variables and the response vector is the dependent variable which is decided on the basis of the independent variables. The Naïve Bayes Classification is done by using the same Bayes Theorem in addition it is assumed that the features/attributes have no correlation, i.e., they are independent of each other and all of the attributes have equal contribution in the classification process and deciding. The Naïve Bayes classifier is of four types; Optimal Naïve Bayes, Gaussian Naïve Bayes, Multinomial Naïve Bayes and Bernoulli Naïve Bayes.

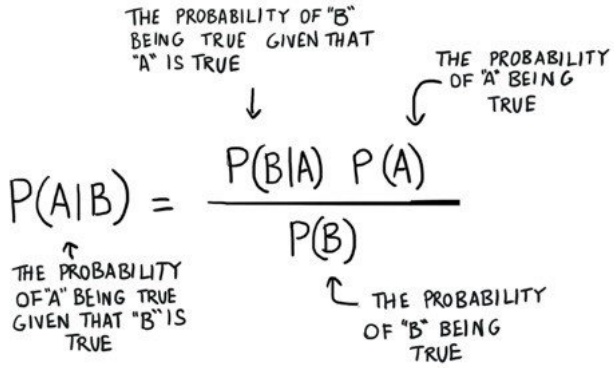

Fig. 3.4 Bayes' Theorem

5. Decision Trees Classification -

Decision tree is a supervised machine learning algorithm that is used for the purpose of both classification and regression but mostly for classification which is

built using the Classification and Regression Tree (CART) algorithm. It consists of a root node, decision nodes, leaf nodes. The decision nodes represent different features that are present in a dataset and the leaf nodes are the output of the decisions. The algorithm is initially trained using a training data set which consists of all the features and attributes along with the values where the algorithm will move when one attribute satisfies the decision node. If the data set has too many decision nodes, the ones that are not needed in the decision making are pruned, i.e., they're removed from the data set. The decision tree can be built on the basis of two values, information gain and gini index. The decision nodes are built as per the information gain calculated for the attributes. Gini index measures the purity/impurity of the data and the decision nodes are built where the gini index is the lowest.

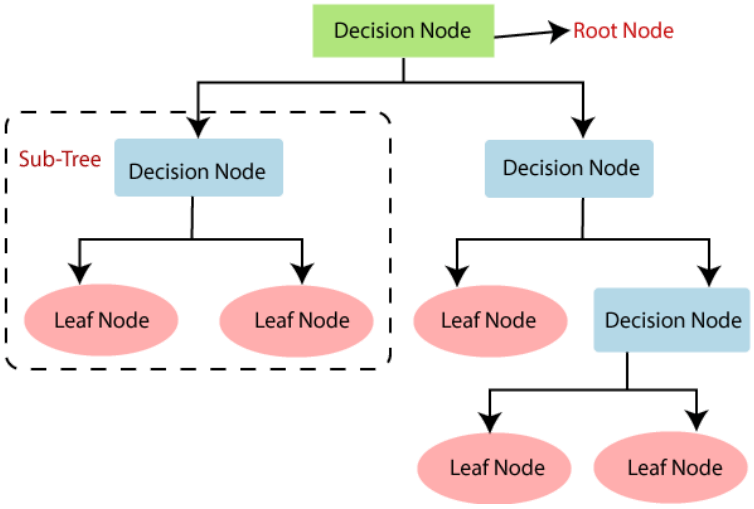

Fig. 3.4 Structure of Basic Decision Tree

6. Support Vector Machine (SVM) -

Support Vector Machine (SVM) is a supervised machine learning algorithm of classifying data points, that is represented on an N-dimensional space and separated by a hyperplane. The objective of an SVM is to find the maximum distance between the two nearest data points having different attributes. Initially

a training data is used to train the algorithm which then classifies the data into two different sets and the decision boundaries, i.e., the hyperplane separates the two different attributes. The nearest data points of different attributes on either side of the hyperplane are known as support vectors and the distance between them is known as margin. These support vectors are the classifiers of the data, i.e., when testing data is analysed through the algorithm the data points will be separated on the basis of these vectors.

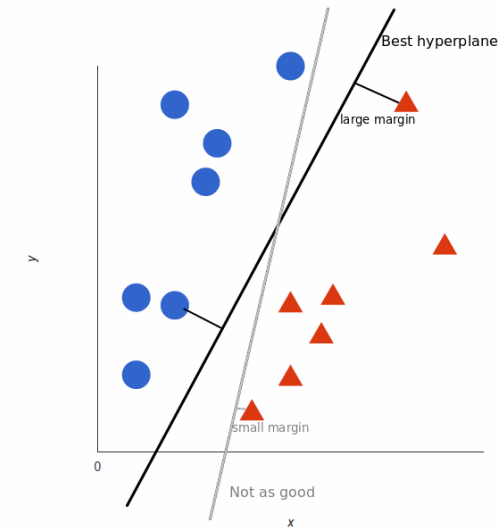

Fig. 3.6 Hyperplane formed using SVM

7. Random Forest -

Random Forest is a supervised machine learning algorithm and as the term suggests, it is a combination of N-number of decision trees that work together using ensemble learning. The random forest algorithm first creates decision trees on the basis of classes while splitting the decision node and combines these decision trees together at the end. After the decision trees have made their predictions which should be having low correlations and then taking an average of the results as some decision trees might predict correctly and some might not. With more number of decision trees, random forest solves the problem of

overfitting in case of large data sets and can predict results with high accuracy with compromising the computing speed.

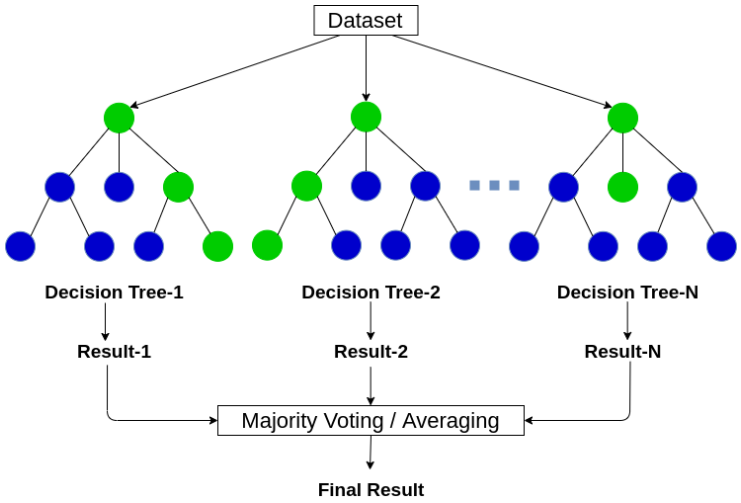

Fig. 3.7 Random Forest Prediction Model

8. Expert Systems -

An expert system is a computer program that behaves and makes decisions like a human brain by simply having a visual perception and understanding of human language for its working. As the term suggests an expert system is an expert in a specific field of study as it is only consisting of domain specific knowledge. An expert system is a high performing and responsive, reliable, and understanding program that can extrapolate the required data from its knowledge base. It consists of three components; a) Knowledge Base - a database that has all the values, rules, knowledge (factual and heuristic) and facts that have been inserted by the human expert; b) Inference Engine - also known as the brain of the expert system as it is able to apply the required rules to the knowledge base in order to arrive at a solution and can follow either forward chaining starting from facts to arrive at solution or backward chaining starting from the solution in order to verify facts; c) User Interface - an interface where the user who is not an expert in a

domain inserts his/her query and the program runs the query through the inference engine and knowledge base and the output is displayed to the user.

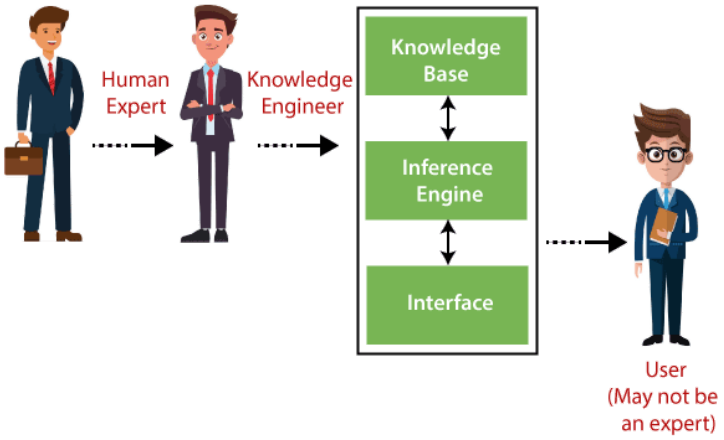

Fig. 3.7 Working of Expert Systems

9. Genetic Algorithms -

Based on Darwin's Theory of Natural Selection which simply puts the statement as survival of the fittest, genetic algorithms are evolutionary algorithms that are used for search and optimization problems in order to have high quality solutions. It consists of N-number of individual solutions known as chromosomes, for a given problem. When a problem arises, the process initiates and all possible solutions/chromosomes are created. A fitness score is assigned to these chromosomes and those individuals that are able to satisfy the fitness score are considered competent for further process. The fittest chromosomes are saughted and the two chromosomes are selected at random and a crossover of their genes takes place in order to have an offspring, i.e., a new solution/chromosome. In order to maintain the diversity of the population mutation is done to some of the features of the offspring which makes the data diverse. This whole process is repeated again and again until the fittest solution is found to the problem.

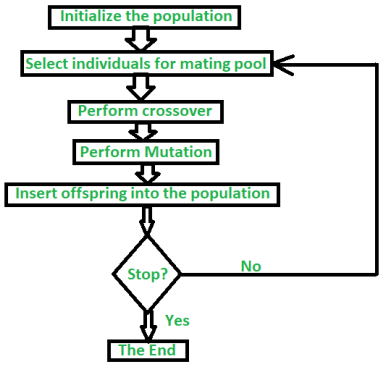

Fig. 3.9 Working of Genetic Algorithm

10. Natural Language Processing (NLP) -

Natural Language Processing is a combination of three fields, namely, Artificial Intelligence, Computer Science and Human Language, i.e., linguistics. As the term suggests NLP algorithms processes the human language in a format that could be easily interpreted by the machines. Text mining is a sufficient method which can derive meaningful information from natural language text, it usually involves the process of structuring the input text, deriving patterns from the structured data and finally evaluating and interpreting the output, which is why text mining and natural language processing goes hand in hand. Natural language processing is divided into two major components; Natural Language Understanding which refers to mapping the given input into natural language into useful representation and analysing those aspects of the language, and Natural Language Generation which is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation. There are various steps involved in natural language processing such as; Tokenization which is the process of breaking the phrases and sentences into small words/tokens, Stemming, i.e., the process of normalising the worlds in their base form, Lemmatization process which converts the base words to the dictionary in order to verify that the word exists, POS (Paths of Speech) Tags indicating the functioning of the words grammatically as well as having a

meaning when used in a sentence, Named Entity Recognition which recognises the proper nouns by using noun phrase identification, phrase classification and entity disambiguation, and lastly Chunking, i.e., the process of combining the tokens and words into chunks or words/phrases. Natural language processing is carried out in five phases; Lexical Analysis, Syntactic Analysis, Semantic Analysis, Discourse Integration and Pragmatic Analysis.

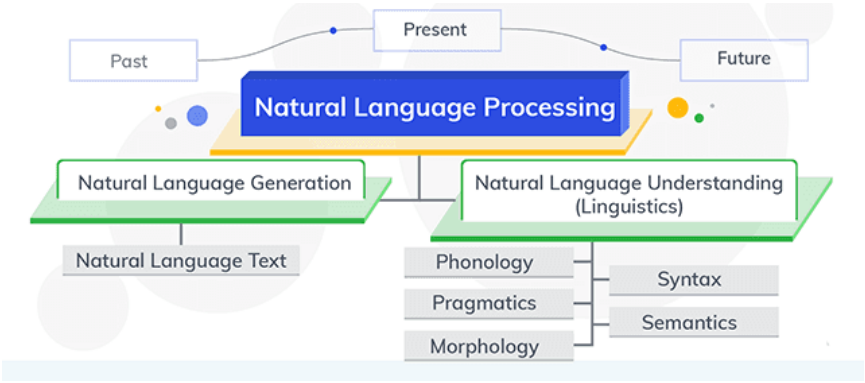

Fig. 3.10 Evolution of NLP

11. Fuzzy Logic -

Fuzzy Logic as the term suggests means not clear, is an extension of the Boolean logic where yes is represented as 1 and no is represented as 0. In Fuzzy Logic the values can lie between 0 and 1 and is given a term accordingly which starts from 0 as absolutely false up till 1 as absolutely true, i.e., partially true and partially false lie in between. The process is divided into four components; rule base - stores the set of IF-THEN rules for which conditions are being given by the human expert for conditional decision making; fuzzification - using the membership function so that crisp numbers that are inputs are converted into fuzzy steps, i.e., LP (Large Positive), MP (Medium Positive), S (Small), MN (Medium Negative), LN (Large Negative); inference engine - the the main component of fuzzy logic that compares the inputs with the IF-THEN rules and decides which values are to be be added with the inputs provided;

defuzzification - converts the values provided by the inference engine into crisp values that could be understood by humans.

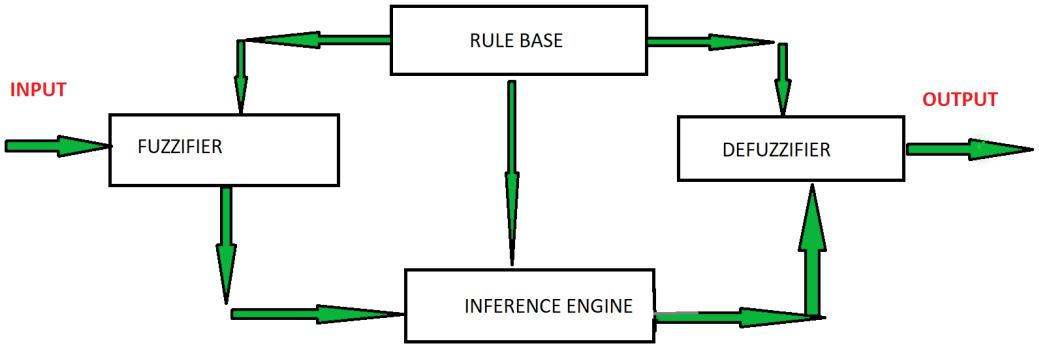

Fig. 3.11 Fuzzy Logic Architecture

12. Association Rule Mining -

Association Rule Mining is a machine learning technique that identifies, learns certain rules in order to apply to large sums of data sets. It is used in order to find patterns/associations in the large data sets in order to find frequently occurring sets. It works on the basis that one thing happens, other will happen or not. The biggest example of this technique is market basket analysis also known as shopping basket analysis which uses algorithms such as Apriori Algorithm in order to find frequently occurring item sets.

Fig. 3.12 Market Basket Analysis using Association Rule Mining

The aim of this study is to apply artificial intelligence in order to improve supply chain management for which Artificial Neural Networks will be the chosen technique since it is the most preferred method due to its high accuracy. A real world Supply Chain Management System (SCMS) delivery data set maintained by the US Agency for International Development (USAID) regarding the supply of medicines and medical equipment for HIV and ARV (Antiretroviral) lab shipments was downloaded from the US Government's data repository. The fields in the data set are similar to Global Fund's PQR (Price, Quality and Reporting) data. The data include deliveries that were fulfilled either via direct drop from vendors or from one of the three Regional Distribution Centers (RDC) in Africa. The data set is imported into python software for analytics and modelling.

Data Analysis

Importing Libraries

Following libraries are being used by the students for data analysis and modelling, and are imported using the import function -

1. NumPy (Data Manipulation) -
   NumPy is a short form for Numerical Python which is used for doing logical and mathematical operations on unidimensional as well as multidimensional array elements and can operate on large data without any complications.

2. Pandas (Data Manipulation) -
   Pandas is a data manipulation and data analysis library in Python. The purpose of this software is to analyse a given time series data. Just like NumPy, Pandas work on multidimensional array elements.

3. Matplotlib (Data Visualization) -

   Matplotlib is the most common data visualization library in Python. With the data provided and command given the library helps in making 2D as well as 3D plots. Other than Python, Matplotlib is available on other programming languages and softwares.

4. Seaborn (Data Visualization) -

   Seaborn is also a data visualization Python library just like Matplotlib. It provides way more informative statistical graphs and plots, and can work with large amounts of data with N-number of variables and is able to work with data frames.

5. SKLearn (Data Pre-processing) -

   SKLearn also known as Scikit-Learn is a data analysis library for statistical modelling and machine learning in Python. It provided supervised learning algorithms such as support vector machines (SVM), decision trees, bayesian classification, etc. and unsupervised learning algorithms such as factor analysis, clustering, principal component analysis (PCA), etc.

6. TensorFlow (Model Creation) -

   TensorFLow is a deep learning software. In Python the TensorFlow library helps in deep learning algorithms, majorly neural networks. It helps in developing and training models for making predictions using APIs like Keras.

Data set and characteristic description

The data set downloaded from Github in the CSV (Comma Separated Values) format, is uploaded from the system in Python using the Pandas library function. It consists of 10324 shipments recorded from June 2006 till November 2014 and consists of 33 variables.

	id	project code	pq #	po / so #	asn/dn #	country	managed by	fulfill via	vendor inco term	shipment mode	pq first sent to client date	po sent to vendor date	scheduled delivery date	delivered to client date	delivery recorded date	product group	sub classification	vendor	
0	1	100-CI-T01	Pre-PQ Process	SCMS-4	ASN-8	Côte d'Ivoire	PMO - US	Direct Drop	EXW	Air	Pre-PQ Process	Date Not Captured	2-Jun-06	2-Jun-06	2-Jun-06	HRDT	HIV test	RANBAXY Fine Chemicals LTD.	
1	3	108-VN-T01	Pre-PQ Process	SCMS-13	ASN-85	Vietnam	PMO - US	Direct Drop	EXW	Air	Pre-PQ Process	Date Not Captured	14-Nov-06	14-Nov-06	14-Nov-06	ARV	Pediatric	Aurobindo Pharma Limited	
2	4	100-CI-T01	Pre-PQ Process	SCMS-20	ASN-14	Côte d'Ivoire	PMO - US	Direct Drop	FCA	Air	Pre-PQ Process	Date Not Captured	27-Aug-06	27-Aug-06	27-Aug-06	HRDT	HIV test	Abbott GmbH & Co. KG	HIV
3	15	108-VN-T01	Pre-PQ Process	SCMS-78	ASN-50	Vietnam	PMO - US	Direct Drop	EXW	Air	Pre-PQ Process	Date Not Captured	1-Sep-06	1-Sep-06	1-Sep-06	ARV	Adult	SUN PHARMACEUTICAL INDUSTRIES LTD (RANBAXY LAB...	
4	16	108-VN-T01	Pre-PQ Process	SCMS-81	ASN-55	Vietnam	PMO - US	Direct Drop	EXW	Air	Pre-PQ Process	Date Not Captured	11-Aug-06	11-Aug-06	11-Aug-06	ARV	Adult	Aurobindo Pharma Limited	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
10319	86818	103-ZW-T30	FPQ-15197	SO-50020	DN-4307	Zimbabwe	PMO - US	From RDC	N/A - From RDC	Truck	10/16/2014	N/A - From RDC	31-Jul-15	15-Jul-15	20-Jul-15	ARV	Pediatric	SCMS from RDC	Lam
10320	86819	104-CI-T30	FPQ-15259	SO-50102	DN-4313	Côte d'Ivoire	PMO - US	From RDC	N/A - From RDC	Truck	10/24/2014	N/A - From RDC	31-Jul-15	6-Aug-15	7-Aug-15	ARV	Adult	SCMS from RDC	
10321	86821	110-ZM-T30	FPQ-14784	SO-49600	DN-4316	Zambia	PMO - US	From RDC	N/A - From RDC	Truck	8/12/2014	N/A - From RDC	31-Aug-15	25-Aug-15	3-Sep-15	ARV	Adult	SCMS from RDC	E
10322	86822	200-ZW-T30	FPQ-16523	SO-51680	DN-4334	Zimbabwe	PMO - US	From RDC	N/A - From RDC	Truck	7/1/2015	N/A - From RDC	9-Sep-15	4-Aug-15	11-Aug-15	ARV	Adult	SCMS from RDC	
10323	86823	103-ZW-T30	FPQ-15197	SO-50022	DN-4336	Zimbabwe	PMO - US	From RDC	N/A - From RDC	Truck	10/16/2014	N/A - From RDC	31-Aug-15	4-Aug-15	11-Aug-15	ARV	Pediatric	SCMS from RDC	La

10324 rows × 33 columns

item description	molecule/test type	brand	dosage	dosage form	unit of measure (per pack)	line item quantity	line item value	pack price	unit price	manufacturing site	first line designation	weight (kilograms)	freight cost (usd)	line item insurance (usd)
HIV, Reveal G3 Rapid HIV-1 Antibody Test, 30 T...	HIV, Reveal G3 Rapid HIV-1 Antibody Test	Reveal	NaN	Test kit	30	19	551.00	29.00	0.97	Ranbaxy Fine Chemicals LTD	True	13	780.34	NaN
Nevirapine 10mg/ml, oral suspension, Bottle, 2...	Nevirapine	Generic	10mg/ml	Oral suspension	240	1000	6200.00	6.20	0.03	Aurobindo Unit III, India	True	358	4521.5	NaN
/ 1/2, Determine Complete HIV Kit, 100 Tests	HIV 1/2, Determine Complete HIV Kit	Determine	NaN	Test kit	100	500	40000.00	80.00	0.80	ABBVIE GmbH & Co.KG Wiesbaden	True	171	1653.78	NaN
Lamivudine 150mg, tablets, 60 Tabs	Lamivudine	Generic	150mg	Tablet	60	31920	127360.80	3.99	0.07	Ranbaxy, Paonta Shahib, India	True	1855	16007.06	NaN
Stavudine 30mg, capsules, 60 Caps	Stavudine	Generic	30mg	Capsule	60	38000	121600.00	3.20	0.05	Aurobindo Unit III, India	True	7590	45450.08	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ivudine/Nevirapine/Zidovudine 30/50/60mg, d...	Lamivudine/Nevirapine/Zidovudine	Generic	30/50/60mg	Chewable/dispersible tablet - FDC	60	166571	599655.60	3.60	0.06	Mylan, H-12 & H-13, India	False	See DN-4307 (ID#:83920)	See DN-4307 (ID#:83920)	705.79
Lamivudine/Zidovudine 150/300mg, tablets, 60 Tabs	Lamivudine/Zidovudine	Generic	150/300mg	Tablet - FDC	60	21072	137389.44	6.52	0.11	Hetero Unit III Hyderabad IN	False	See DN-4313 (ID#:83921)	See DN-4313 (ID#:83921)	161.71
Efavirenz/Lamivudine/Tenofovir Disoproxil Fuma...	Efavirenz/Lamivudine/Tenofovir Disoproxil Fuma...	Generic	600/300/300mg	Tablet - FDC	30	514526	5140114.74	9.99	0.33	Cipla Ltd A-42 MIDC Mahar. IN	False	Weight Captured Separately	Freight Included in Commodity Cost	5284.04
Lamivudine/Zidovudine 150/300mg, tablets, 60 Tabs	Lamivudine/Zidovudine	Generic	150/300mg	Tablet - FDC	60	17465	113871.80	6.52	0.11	Mylan (formerly Matrix) Nashik	True	1392	Freight Included in Commodity Cost	134.03
mivudine/Zidovudine 30/60mg, dispersible tab...	Lamivudine/Zidovudine	Generic	30/60mg	Chewable/dispersible tablet - FDC	60	36639	72911.61	1.99	0.03	Cipla, Goa, India	False	Weight Captured Separately	Freight Included in Commodity Cost	85.82

Fig. 3.13 Shipment dataset in Python

Data set description -

- id - Identification number
- project code - Project Code

- pq # - Price Quote number
- po / so # - Purchase order/Supply order
- asn/dn # - Advanced Shipment Note for direct drop/for Regional Distribution Centres deliveries
- country - Destination Country
- managed by - SCMS managing office: either the Program Management Office (PMO) in the U.S. or the relevant SCMS field office
- fulfill via - Method through which the shipment was fulfilled: via Direct Drop from vendor or from stock available in the RDCs
- vendor incoterm - The vendor INCOterm for Direct Drop deliveries
- shipment mode - Method by which commodities are shipped
- pq first sent to client date - Date the PQ is first sent to the client
- po sent to vendor date - Date the PO is first sent to the vendor
- scheduled delivery date - Current anticipated delivery date
- delivered to client date - Date of delivery to client
- delivery recorded date - Date on which delivery to client was recorded in SCMS information systems
- product group - Product group for item
- sub classification - Identifies relevant product sub classifications
- vendor - Vendor name
- item description - Product name and formulation from Partnership for Supply Chain Management (PFSCM) Item Master
- molecule/test type - Active drug(s) or test kit type
- brand - Generic or branded name for the item
- dosage - Item dosage and unit
- dosage form - Dosage form for the item (tablet, oral solution, injection, etc.)
- unit of measure (per pack) - Pack quantity (pills or test kits) used to compute unit price
- line item quantity - Total quantity (packs) of commodity per line item

- line item value - Total value of commodity per line item
- pack price - Cost per pack
- unit price - Cost per pill or per test
- manufacturing site - Identifies manufacturing site for the line item for direct drop and from RDC deliveries
- first line designation - Designates if the line in question shows the aggregated freight costs and weight associated with all items on the ASN/DN
- weight (kilograms) - Weight for all lines on an ASN/DN
- freight cost (usd) - Freight charges associated with all lines on the respective ASN/DN
- line item insurance (usd) - Line item cost of insurance

Variables and their Characteristics -

- id Integer
- project code Object
- pq # Object
- po / so # Object
- asn/dn # Object
- country Object
- managed by Object
- fulfill via Object
- vendor incoterm Object
- shipment mode Object
- pq first sent to client date Object
- po sent to vendor date Object
- scheduled delivery date Object
- delivered to client date Object
- delivery recorded date Object

- product group Object
- sub classification Object
- vendor Object
- item description Object
- molecule/test type Object
- brand Object
- dosage Object
- dosage form Object
- unit of measure (per pack) Integer
- line item quantity Integer
- line item value Float
- pack price Float
- unit price Float
- manufacturing site Object
- first line designation Bool
- weight (kilograms) Object
- freight cost (usd) Object
- line item insurance (usd Float

Data Preprocessing

When it comes to real world data sets, inaccurate, inconsistent and incomplete data are the most common problems that are faced while doing analysis and drawing inferences. In order to create perfect machine learning algorithms, accuracy, consistency and completeness are the factors that determine the quality of the data. Data preprocessing is the process of converting the raw real world data ready for applying machine learning algorithms and is very crucial in model creation. This only helps in increasing the efficiency and accuracy of the model created. The tasks involved in data preprocessing are -

1. Data Cleaning -

   In this step, the missing values in the data set are checked, then it is upto the person conducting the study what to do with the missing values such as ignoring those rows/tuples, using a measure of central tendency (mean or mode) or most probable value for filling the missing values. Sometimes there is a variance or random error in the data set known as noisy data. These noisy data can be replaced by using the process of binning where the values from neighbour data points are taken into consideration such as smoothing by means, medians and even boundaries.

2. Data Integration -

   Sometimes it is required to merge data from multiple sources, meaning multiple data sets being integrated together but this process has to be done carefully in order to avoid inconsistencies and redundancies in the new data set. In order to avoid these situations measures such as entity identification, correlation analysis such as chi-squared test, correlation coefficient and covariance calculation for nominal and numeric data respectively are done. It helps in improving the accuracy and speed of further data analysis.

3. Data Reduction -

   In this step the original data set is replaced by a smaller version of itself made of values selected from it through the process of dimensionality reduction techniques such as principal component analysis, attribute subset selection and wavelet transformation for reducing the number of random variable, and numerosity reduction techniques such as parametric models like regression and log-linear models and non-parametric models like clustering, data aggregation, sampling and histograms for replacing the original data set with alternative smaller data set.

4. Data Transformation -

   The data is converted into appropriate form for mining purposes by using techniques such as normalization. It also involves various data discretization techniques like binning, cluster analysis, histogram analysis, correlation analysis and decision tree analysis. Data transformation can also be done using the concept hierarchy generation based on distinct values as well as schema definitions per attribute.

Checking for Non-null values

In order to have an accurate prediction model, it is really necessary for the data to be clean and not contain any empty values, i.e., they should be non-null values. By simply inserting the command “name of data frame” and “.info()” the person can get the details such as what data type is each column/variable and the count of all the non-null values in each variable. In Fig. 3.14 It can be observed that one of the variables, “first line designation” has a data type of “bool”, i.e., boolean values of true and false. For further process it is necessary that those are in “object” type. It can be done by changing the data type using the command “name of data frame[name of variable]” and “.astype(str)” which will change the data type of the variable to “object”.

In case the person wants to know the total number of NA values in each column they can simply insert the command “name of data frame” and “.isna().sum()” which returns the sum of NA values for each variable.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10324 entries, 0 to 10323
Data columns (total 33 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   id                             10324 non-null  int64
 1   project code                   10324 non-null  object
 2   pq #                           10324 non-null  object
 3   po / so #                      10324 non-null  object
 4   asn/dn #                       10324 non-null  object
 5   country                        10324 non-null  object
 6   managed by                     10324 non-null  object
 7   fulfill via                    10324 non-null  object
 8   vendor inco term               10324 non-null  object
 9   shipment mode                  9964 non-null   object
 10  pq first sent to client date   10324 non-null  object
 11  po sent to vendor date         10324 non-null  object
 12  scheduled delivery date        10324 non-null  object
 13  delivered to client date       10324 non-null  object
 14  delivery recorded date         10324 non-null  object
 15  product group                  10324 non-null  object
 16  sub classification             10324 non-null  object
 17  vendor                         10324 non-null  object
 18  item description               10324 non-null  object
 19  molecule/test type             10324 non-null  object
 20  brand                          10324 non-null  object
 21  dosage                         8588 non-null   object
 22  dosage form                    10324 non-null  object
 23  unit of measure (per pack)     10324 non-null  int64
 24  line item quantity             10324 non-null  int64
 25  line item value                10324 non-null  float64
 26  pack price                     10324 non-null  float64
 27  unit price                     10324 non-null  float64
 28  manufacturing site             10324 non-null  object
 29  first line designation         10324 non-null  bool
 30  weight (kilograms)             10324 non-null  object
 31  freight cost (usd)             10324 non-null  object
 32  line item insurance (usd)      10037 non-null  float64
dtypes: bool(1), float64(4), int64(3), object(25)
memory usage: 2.5+ MB
```

Fig. 3.14 Checking for non-null values

```
ID                                0
Project Code                      0
PQ #                              0
PO / SO #                         0
ASN/DN #                          0
Country                           0
Managed By                        0
Fulfill Via                       0
Vendor INCO Term                  0
Shipment Mode                   360
PQ First Sent to Client Date      0
PO Sent to Vendor Date            0
Scheduled Delivery Date           0
Delivered to Client Date          0
Delivery Recorded Date            0
Product Group                     0
Sub Classification                0
Vendor                            0
Item Description                  0
Molecule/Test Type                0
Brand                             0
Dosage                         1736
Dosage Form                       0
Unit of Measure (Per Pack)        0
Line Item Quantity                0
Line Item Value                   0
Pack Price                        0
Unit Price                        0
Manufacturing Site                0
First Line Designation            0
Weight (Kilograms)                0
Freight Cost (USD)                0
Line Item Insurance (USD)       287
dtype: int64
```

Fig. 3.15 Checking for NA values

It can be seen from Fig. 3.14 and Fig. 3.15 that out of the 33 variables/columns, 30 columns have the same number of non-null values as the total number of records in the data set, i.e., 10324 records, and 3 columns have NA values which are “shipment mode”, “dosage” and “line item insurance (usd)”.

For data analysis it is important to know whether the columns that have NA values in them should be dropped on the basis of percentage of missing values in them or those NA values could be replaced by either mean, i.e., average of all the values or mode, i.e., the most frequently occurring value in the column. It can be checked by inserting the command “name of data frame” and “.isna().mean()” in Python.

```
id                              0.000000
project code                    0.000000
pq #                            0.000000
po / so #                       0.000000
asn/dn #                        0.000000
country                         0.000000
managed by                      0.000000
fulfill via                     0.000000
vendor inco term                0.000000
shipment mode                   0.034870
pq first sent to client date    0.000000
po sent to vendor date          0.000000
scheduled delivery date         0.000000
delivered to client date        0.000000
delivery recorded date          0.000000
product group                   0.000000
sub classification              0.000000
vendor                          0.000000
item description                0.000000
molecule/test type              0.000000
brand                           0.000000
dosage                          0.168152
dosage form                     0.000000
unit of measure (per pack)      0.000000
line item quantity              0.000000
line item value                 0.000000
pack price                      0.000000
unit price                      0.000000
manufacturing site              0.000000
first line designation          0.000000
weight (kilograms)              0.000000
freight cost (usd)              0.000000
line item insurance (usd)       0.027799
dtype: float64
```

Fig. 3.16 Percentage of missing values

In Fig.3.16, it can be seen that the three columns “shipment mode”, “dosage” and “line item insurance (usd)” have 3.48%, 16.81% and 2.78% of missing values respectively.

In this study with the SCMS delivery data set, all the unwanted variables will be removed and the treatment of NA values will be decided as work progresses.

The “id” column is not required as each row/tuple in this column consists of a unique value and in model creation this will not serve any purpose and will not be useful in the prediction.

Ideally when it comes to missing values in model creation, if the variable consists of more than 23-25% missing values, the column should be dropped in order to maintain the efficiency and accuracy of the model. Out of the three variables, “shipment mode”,

“dosage” and “line item insurance (usd)”, none consists of more than 25% missing values, hence no column shall be dropped.

Since “shipment mode” is the target variable, i.e., the values for this variable will be predicted through the created model, the tuples/rows consisting of missing values will be dropped which results in leaving 9964 tuples/rows in the dataset with each tuple/row in the “shipment mode” column consisting of a value. For the “dosage” and “line item insurance (usd)” columns, the missing values can be replaced by using a measure of central tendency, i.e., mean, median or mode. In the “dosage” column, there are many tuples that consist of multiple values, meaning that mean cannot be used as it works on tuples containing single values, therefore, in this case mode will be used in order to replace the missing values. The “line item insurance (usd) column consists of numeric values and each tuple consists of a single value, therefore, mean is the preferred method of replacing the missing values.

The data columns, i.e., “pq first sent to client date”, “po sent to vendor date”, “scheduled delivery date”, “delivered to client date” and “delivery recorded date” should be checked in order to find whether there are any values that are not in accordance with the standard date time format. If there are any improper values, they wouldn’t be shown as missing values by Python. In order to do that while checking for these improper date time values, coercing of the errors can be done and the improper values will be converted into “NaT”, i.e., not a datetime value. These NaT values are the new missing values in the data set and it should be checked whether these missing values should be replaced or the whole column should be dropped.

```
pq first sent to client date 0.23293857888398234
po sent to vendor date 0.5436571657968687
scheduled delivery date 0.0
delivered to client date 0.0
delivery recorded date 0.0
```

Fig. 3.17 Percentage of missing values in date columns

From Fig. 3.17, it can be seen that two columns, “pq first sent to client date” and “po sent to vendor date” have 23.2% and 54.3% missing values meaning that these columns should be dropped in order to maintain the efficiency of the data set. In order to have proper data analysis, the remaining three date columns should be extracted, i.e., the values of day, month and year should be extracted into separate columns and the original columns should be dropped in order to avoid data redundancy.

scheduled delivery date Year	scheduled delivery date Month	scheduled delivery date Day	delivered to client date Year	delivered to client date Month	delivered to client date Day	delivery recorded date Year	delivery recorded date Month	delivery recorded date Day
2006	6	2	2006	6	2	2006	6	2
2006	11	14	2006	11	14	2006	11	14
2006	8	27	2006	8	27	2006	8	27
2006	9	1	2006	9	1	2006	9	1
2006	8	11	2006	8	11	2006	8	11
...	...	...	...	...	...	...	...	...
2015	7	31	2015	7	15	2015	7	20
2015	7	31	2015	8	6	2015	8	7
2015	8	31	2015	8	25	2015	9	3
2015	9	9	2015	8	4	2015	8	11
2015	8	31	2015	8	4	2015	8	11

Fig. 3.18 Extracted date columns

It can be observed that two columns, “weight (kilograms)” and “freight cost (usd)” have improper values in them, such as Weight Captured Separately and See DN-4313 (ID#:83921) in “weight (kilogram)” column and Freight Included in Commodity Cost and See DN-4313 (ID#:83921) in the “freight cost (usd)” column. These values are converted into missing values. The mean of missing values is calculated for both the columns and it is observed that “weight (kilogram) has 38.1% missing values whereas “freight cost (usd)” has 96.1% missing values, meaning both the columns should be dropped in order to avoid any discrepancies.

The remaining object columns are checked in order to find the total number of unique values in them. This is done in order to find the high-cardinality columns. High-cardinality columns/variables are those that have way too many unique/uncommon values which might slow down the performance of the created model substantially.

```
{'asn/dn #': 6804,
 'brand': 47,
 'country': 43,
 'dosage': 54,
 'dosage form': 17,
 'first line designation': 2,
 'fulfill via': 2,
 'item description': 183,
 'managed by': 4,
 'manufacturing site': 88,
 'molecule/test type': 85,
 'po / so #': 6013,
 'pq #': 1237,
 'product group': 5,
 'project code': 142,
 'shipment mode': 4,
 'sub classification': 6,
 'vendor': 72,
 'vendor inco term': 8}
```

Fig. 3.19 Unique values in each object column

From Fig. 3.19, it can be seen that three columns, namely, “asn/dn #”, “po / so #” and “pq #” have unique values of count 6804, 6013 and 1237 respectively. These columns are high-cardinality columns and should be avoided when dealing with time series data as they will slow down the performance of the model. So the three columns shall be dropped.

In Fig. 3.19, it can also be seen that two columns, “fulfill via” and “first line designation” both have only 2 unique values, ‘Direct drop’ and ‘From RDC’ in “fulfill via” column and ‘True’ and ‘False’ in “first line designation” column. For modelling purposes, binary encoding can be done on these columns, i.e., use of 0 and 1. This will change the values in the two columns to 0s and 1s depending upon which value is assigned 0 and which is assigned 1.

For the remaining object columns, it is necessary for them to be converted into numerical format as well for model creation and data analysis. For this purpose, one-hot encoding is done, which extracts each value in each column and creates a separate column for that value. It makes the data more expressive by converting the categorical data into numerical format. Since “shipment mode” is the target column, one-hot encoding shall not be applied to it and will be dealt with later in the process. On the remaining object columns, one-hot encoding is done by extracting dummy variables of each variable, concatenating these dummy variables with the data set and dropping the original columns.

The data set is divided into “x”, being the input and “y” being the output, which is “shipment mode”. The “shipment mode” column consists of object data type, and since TensorFlow is being used for model prediction, it is converted into numerical format by label mapping, i.e., all four values in the “y” data set are assigned a unique number.

In “x” all the columns have different ranges of values which will be a problem in modelling thus they are scaled down having the same range of values using the StandardScaler() function which gives each column a mean of zero and variance of 1.

	fulfill via	unit of measure (per pack)	line item quantity	line item value	pack price	unit price	first line designation	line item insurance (usd)	scheduled delivery date Year	scheduled delivery date Month	scheduled delivery date Day	delivered to client date Year	delivered to client date Month	delivered to client date Day	delivery recorded date Year	delivery recorded date Month	delivery recorded date Day	
0	-1.022329	-0.620836	-0.458308	-0.454432	0.158961	0.165632	0.681493	-1.737426e-15	-2.248450	-0.150212	-1.677498	-2.235451	-0.143479	-1.608361	-2.233013	-0.153924	-1.608836	-0
1	-1.022329	2.128210	-0.434128	-0.438261	-0.339861	-0.244388	0.681493	-1.737426e-15	-2.248450	1.370524	-0.401388	-2.235451	1.359492	-0.243784	-2.233013	1.346170	-0.238576	-0
2	-1.022329	0.295513	-0.446452	-0.341503	1.274746	0.091479	0.681493	-1.737426e-15	-2.248450	0.458083	0.981064	-2.235451	0.457709	1.234508	-2.233013	0.446114	1.245872	-0
3	-1.022329	-0.228115	0.327996	-0.091418	-0.388212	-0.226940	0.681493	-1.737426e-15	-2.248450	0.762230	-1.783840	-2.235451	0.758303	-1.722075	-2.233013	0.746132	-1.723024	-0
4	-1.022329	-0.228115	0.477858	-0.107909	-0.405496	-0.235664	0.681493	-1.737426e-15	-2.248450	0.458083	-0.720416	-2.235451	0.457709	-0.584928	-2.233013	0.446114	-0.581141	-0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
9959	0.978159	-0.228115	3.646911	1.260606	-0.396744	-0.231302	-1.467367	9.311403e-01	1.497546	0.153936	1.406433	1.496516	0.157115	-0.130069	1.491209	0.146095	0.446554	-0
9960	0.978159	-0.228115	0.060612	-0.062709	-0.332860	-0.209492	-1.467367	-1.594184e-01	1.497546	0.153936	1.406433	1.496516	0.457709	-1.153502	1.491209	0.446114	-1.037894	-0
9961	0.978159	-0.620836	12.223401	14.258435	-0.256943	-0.113530	-1.467367	1.010782e+01	1.497546	0.458083	1.406433	1.496516	0.457709	1.007078	1.491209	0.746132	-1.494648	-0
9962	0.978159	-0.228115	-0.028294	-0.130032	-0.332860	-0.209492	0.681493	-2.149004e-01	1.497546	0.762230	-0.933101	1.496516	0.457709	-1.380931	1.491209	0.446114	-0.581141	-0
9963	0.978159	-0.228115	0.444312	-0.247288	-0.431968	-0.244388	-1.467367	-3.115330e-01	1.497546	0.458083	1.406433	1.496516	0.457709	-1.380931	1.491209	0.446114	-0.581141	-0

Fig. 3.20 Normalised values in data set “x”

This completes the data pre-processing stage.

Model Creation

Now the whole data preprocessing is done, the next stage is model creation. In order to create a prediction model and for checking its accuracy, it is necessary to have both training data as well as testing data. In this case the data will be divided into training and testing data in a 70-30 ratio using the train_test_split() function. Both “x” and “y” datasets will be divided into “x_train”, “x_test”, “y_train” and “y_test”. The data will be shuffled in order to avoid any inefficiencies while testing. This splits the data into 6974 tuples for training the model and 2990 tuples for testing the model for both “x” and “y”.

The input layer is created using the Keras library for neural networks that will be using TensorFlow in the backend for model creation. Hidden/dense layers are used with “ReLU” activation function that converts inputs into outputs if and only if the value is positive or else it will be zero. The output layer will be having the “softmax” activation function so that each of the outputs will have probabilities between 0 to 1 and having a total sum of 1. For early stopping callbacks, the patience is set to three, meaning if after three consecutive epochs the validation loss of the model is not improving, the model will return the best weights from the best epoch.

Now the model is trained and ready for prediction using the test data.

4. RESULTS

During the course of this project, the students gained a lot of insights about data science, machine learning and artificial intelligence. With the help of their project guide, Dr. Kamal Gulati, the students were able to complete the project. They learned a lot about supply chain management, Route to Market (RtM), SCOR Model, incoterms, various machine learning algorithms and how artificial intelligence can help in optimizing the supply chain by applying it on real world data sets. They learned through continuous communication with their project guide how necessary and valuable it is to have a continuous flow of information and idea sharing among a student and his guide that will only help in efficient working of the student. Along with this the students were able to gain hands-on experience with different libraries used in Python programming language throughout the course of this project and how they were able to understand a real world data set, its characteristics and how the data should be treated in order to create a prediction model. The students believe that this experience will make their future career path more smooth and easy.

The purpose of this study was creation of a prediction model for predicting the mode of shipment for transfer of medicines and medical equipment for HIV and ARV based on the dataset maintained by the US Agency of International Development (USAID).

With the primary objective of the study, i.e., completion of the project being completed by creating the prediction model using Artificial Neural Networks (ANN) and after training and testing the model the following results derived can be interpreted in order to have a clear understanding of what happened.

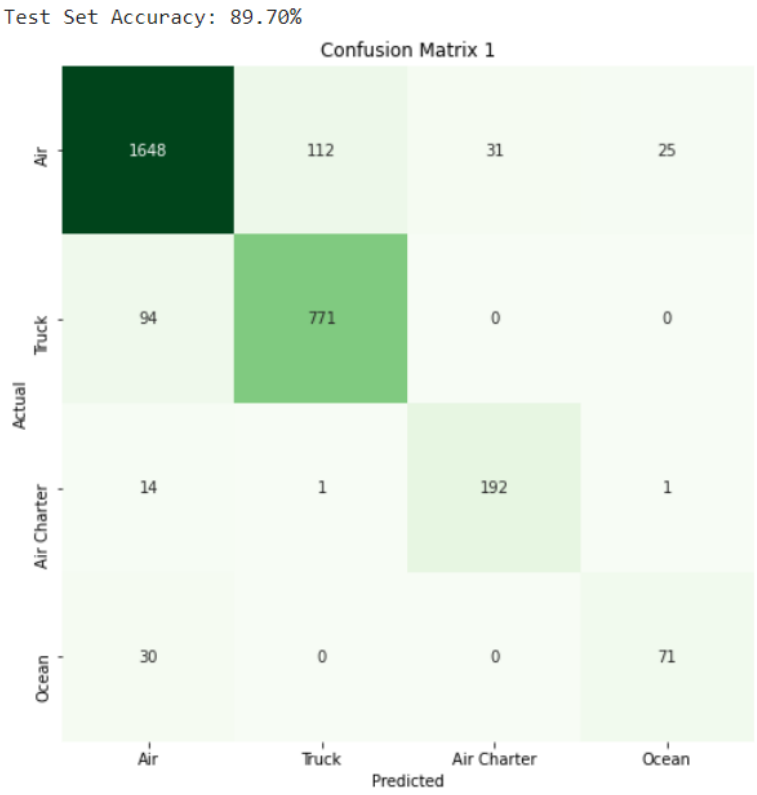

Fig. 4.1 Confusion Matrix of predicted shipment mode

In the Fig. 4.1, the confusion matrix created shows the shipment modes predicted by the neural networks model as well as those ones where the predictions came true. It can be seen that “Air” mode had the maximum number of predictions true, i.e., out of all the times when the mode of shipment was predicted by model as “Air” mode, 1684 times the prediction was correct.

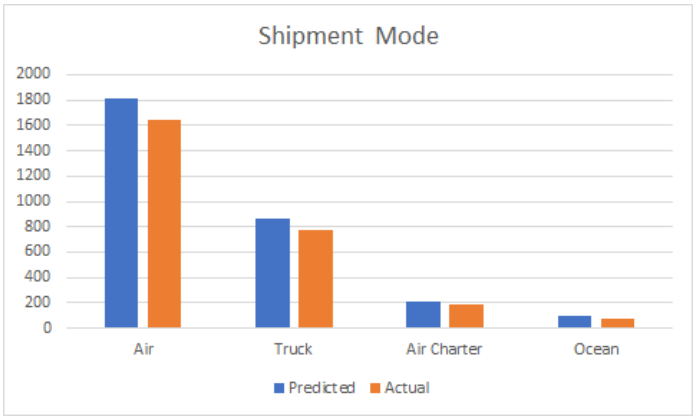

Fig. 4.2 Comparison of predicted vs actual occurrences of shipment modes

In Fig. 4.2 It can be seen that after training the model and using the test data for predicting the shipment mode, “Air” mode is the most preferred mode for transferring the medicines and medical equipment. After that comes “Truck”, “Air Charter” and “Ocean” respectively.

Classification Report 1:

	precision	recall	f1-score	support
Air	0.92	0.91	0.92	1816
Truck	0.87	0.89	0.88	865
Air Charter	0.86	0.92	0.89	208
Ocean	0.73	0.70	0.72	101
accuracy			0.90	2990
macro avg	0.85	0.86	0.85	2990
weighted avg	0.90	0.90	0.90	2990

Fig. 4.3 Classification Report of predicted shipment mode

Precision in simple words means the number of true positives of a variable divided by the total number of positives/predictions that came out to be true. Its formula is -

$$\frac{\text{True Positives (TP)}}{\text{True Positives (TP) + False Positives (FP)}}$$

From the classification report generated for the prediction model, we can see that “Air” mode has the highest precision with 92% whereas “Ocean” has the lowest precision with only 73%.

Recall in simple words means the number of times true positives were identified correctly by the model, i.e., the number of times the model predicted correctly. Its formula is -

$$\frac{\text{True Positives (TP)}}{\text{True Positives (TP) + False Negatives (FN)}}$$

From the classification report generated for the prediction model, we can see that “Air Charter” has the highest recall with 92%, i.e., out of all the predictions made for “Air Charter” as the shipment mode, 92% were correct, whereas “Ocean” has the lowest recall with only 70% predictions correct.

Accuracy is the total number of correct predictions made by the model divided by the total number of predictions for all the classes made by the model. It is calculated in order to know how often the algorithm makes a classification correct. Its formula is -

$$\frac{\text{True Positives + True Negatives}}{\text{True Positives + False Positives + True Negatives + False Negatives}}$$

From the classification report, it can be seen that the Artificial Neural Networks (ANN) model created for predicting the shipment mode has an **accuracy of 90%**.

5. FINDINGS AND RECOMMENDATIONS

In this study, the use of the famous machine learning algorithm of artificial intelligence, Artificial Neural Networks (ANN) was used in order to create a model that will predict the mode of shipment for HIV and ARV medicines and medical equipment based on the information provided by the user.

The data set, originally having 10324 records of shipments, shows that between the years 2006 and 2014, vast majority of the shipments were having African destinations, countries such as South Africa, Nigeria, Côte d'Ivoire, etc. accounted for 83.91% of the shipments because African countries have been in major need of HIV treatment and ARV (Antiretroviral) Therapy since 2006. The remaining shipments were sent to countries like Pakistan, Vietnam, Afghanistan, etc.

The artificial neural networks (ANN) model created suggests that most of the shipments were transported through air, i.e., 60%, whereas the sea route was not the preferred mode for the shipment by majority of the vendors with not even 5% preferring for the ocean route.

Applying neural networks to inspect medical shipments by means of thinking about historical data, dosage of medicines, the country they’re being sent to, etc. is really helpful for more than one stakeholder. In a dynamic situation like a supply chain for shipments, there is difficulty in predicting the perfect outcomes. In this study, Artificial Neural Networks (ANN) were applied to the real world data set for predicting the shipment mode and after creating a confusion matrix after applying the mentioned algorithm gives an accuracy of 90% in predicting the shipment mode to be used. With the help of ANN and other machine learning algorithms for artificial intelligence, the supply chain management can be optimized over all areas.

There are a few recommendations that can be made for the future studies -

- Getting access to more data for improving the efficiency of the model.
- Name of the shipment sending country would be highly helpful as it will help in determining which country sends most shipments to which country.
- In situations like COVID-19 pandemic, transport of necessary medicines and medical equipment can be studied as there are situations when one country is sender at one instance and on the receiving end at another instance.
- Creating prediction models using other major algorithms such as -
    - Fuzzy Logic
    - Genetic Algorithms
    - Decision Trees
    - Association Rule Mining
    - Naïve Bayes Classification
    - Support Vector Machines

6. LIMITATIONS OF THE STUDY

As limitations are the mandatory part of any project, this one is also not an exception to this. This study consists of a few limitations, such as -

- The collection of data was limited as the US Agency for International Development even after updating the metadata on 6th April 2021 uploaded the data between 2006 and 2014.
- Restricted access to various artificial intelligence in supply chain management research papers.
- The quality of the result entirely depends upon the quality of data.
- Initial data set had way too many improper and missing values.
- Using the simplest machine learning algorithms might result in creating more accurate models, but they can't be trusted as in order to avoid overfitting only a few variables will be taken into account.

With all the limitations, the students tried their best to make this report as best as possible.

7. BIBLIOGRAPHY

1. Min, H. (2010). Artificial intelligence in supply chain management: theory and applications. International Journal of Logistics: Research and Applications, 13(1), 13-39.
2. Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., & Fischl, M. (2021). Artificial intelligence in supply chain management: A systematic literature review. Journal of Business Research, 122, 502-517.
3. Liu, H. (2015, April). Forecasting Model of Supply Chain Management Based on Neural Network. In 2015 International Conference on Automation, Mechanical Control and Computational Engineering (pp. 179-183). Atlantis Press.
4. Ciupan, E. (2014). A study regarding the possibility of optimizing the supply batch using artificial neural networks. Procedia Engineering, 69, 141-149.
5. Silva, N., Ferreira, L. M. D., Silva, C., Magalhães, V., & Neto, P. (2017). Improving supply chain visibility with artificial neural networks. Procedia Manufacturing, 11, 2083-2090.

https://www.enchange.com/
https://catalog.data.gov/dataset/supply-chain-shipment-pricing-data
https://ribble-pack.co.uk/blog/supply-chain-examples
https://www.supplychaindigital.com/supply-chain-2/ibm-digitally-perfecting-supply-chain
https://en.wikipedia.org/wiki/Supply_chain
https://en.wikipedia.org/wiki/Supply_chain_management
https://en.wikipedia.org/wiki/Incoterms
https://en.wikipedia.org/wiki/Supply_chain_operations_reference
https://www.sciencedirect.com/topics/computer-science/rough-set-theory

https://www.javatpoint.com/machine-learning-random-forest-algorithm
https://www.javatpoint.com/expert-systems-in-artificial-intelligence
https://www.geeksforgeeks.org/genetic-algorithms/
https://en.wikipedia.org/wiki/Natural_language_processing
https://www.javatpoint.com/fuzzy-logic

8. ANNEXURE

South Africa	1406	Kenya	111	Lebanon	8
Nigeria	1194	Burundi	98	Angola	7
Côte d'Ivoire	1083	Namibia	95	Liberia	6
Uganda	779	Cameroon	75	Lesotho	4
Vietnam	688	Botswana	70	Sierra Leone	4
Zambia	683	Ghana	58	Afghanistan	3
Haiti	655	Dominican Republic	52	Togo	3
Mozambique	631	Sudan	46	Senegal	3
Zimbabwe	538	Swaziland	35	Kyrgyzstan	2
Tanzania	519	Mali	17	Kazakhstan	2
Rwanda	430	Pakistan	15	Burkina Faso	2
Congo, DRC	333	Guatemala	15	Belize	1
Guyana	237	Malawi	14	Guinea	1
Ethiopia	216	Benin	13		
South Sudan	164	Libya	8		

Fig. 8.1 Destination countries for the shipments in descending order

```
PMO - US                      10265
South Africa Field Office        57
Ethiopia Field Office             1
Haiti Field Office                1
Name: managed by, dtype: int64
```

Fig. 8.2 SCMS Managing office

```
Air            6113
Truck          2830
Air Charter     650
Ocean           371
Name: shipment mode, dtype: int64
```

Fig. 8.3 Shipment mode used

```
ARV     8550
HRDT    1728
ANTM      22
ACT       16
MRDT       8
Name: product group, dtype: int64
```

Fig. 8.4 Product Group of items

```
Adult                  6595
Pediatric              1955
HIV test               1567
HIV test - Ancillary    161
Malaria                  30
ACT                      16
Name: sub classification, dtype: int64
```

Fig. 8.5 Sub classification of items

```
N/A - From RDC    5404
EXW               2778
DDP               1443
FCA                397
CIP                275
DDU                 15
DAP                  9
CIF                  3
Name: vendor inco term, dtype: int64
```

Fig. 8.6 Incoterms followed by the Vendor

Generic	7285	Isentress	44	Coartem	12
Determine	799	Prezista	42	Viracept	11
Uni-Gold	373	Epivir	42	Multispot	5
Aluvia	250	Videx EC	41	DoubleCheck	5
Kaletra	165	Retrovir	41	INSTi	5
Norvir	136	Ziagen	37	Paramax	5
Stat-Pak	115	Crixivan	36	LAV	4
Bioline	113	Capillus	35	Reveal	3
Truvada	94	Intelence	32	ImmunoComb	3
Videx	84	Genie	30	Combivir	3
Colloidal Gold	70	Viramune	28	Hexagon	3
Stocrin/Sustiva	69	Clearview	19	InstantCHEK	2
OraQuick	60	Reyataz	18	Bundi	2
Invirase	53	Trizivir	18	CareStart	1
Viread	52	Atripla	16	Visitect	1
Zerit	46	First Response	15	Pepti-LAV	1

Fig. 8.7 Brands of medicines and medical equipments shipped

```
Tablet                                3532
Tablet - FDC                          2749
Test kit                              1575
Capsule                                729
Oral solution                          727
Chewable/dispersible tablet - FDC      239
Oral suspension                        214
Test kit - Ancillary                   161
Chewable/dispersible tablet            146
Delayed-release capsules               131
Delayed-release capsules - blister      41
Powder for oral solution                28
Tablet - FDC + co-blister               20
Tablet - FDC + blister                  15
Tablet - blister                        10
Injection                                6
Oral powder                              1
Name: dosage form, dtype: int64
```

Fig. 8.8 Form of dosage

```
Epoch 1/100
175/175 [==============================] - 2s 6ms/step - loss: 0.6241 - accuracy: 0.7701 - val_loss: 0.3372 - val_accuracy: 0.8832
Epoch 2/100
175/175 [==============================] - 1s 3ms/step - loss: 0.2296 - accuracy: 0.9019 - val_loss: 0.2912 - val_accuracy: 0.8975
Epoch 3/100
175/175 [==============================] - 1s 3ms/step - loss: 0.1903 - accuracy: 0.9199 - val_loss: 0.2949 - val_accuracy: 0.8889
Epoch 4/100
175/175 [==============================] - 1s 3ms/step - loss: 0.1698 - accuracy: 0.9220 - val_loss: 0.2867 - val_accuracy: 0.8910
Epoch 5/100
175/175 [==============================] - 1s 3ms/step - loss: 0.1661 - accuracy: 0.9258 - val_loss: 0.2873 - val_accuracy: 0.8989
Epoch 6/100
175/175 [==============================] - 1s 3ms/step - loss: 0.1518 - accuracy: 0.9367 - val_loss: 0.2908 - val_accuracy: 0.8939
Epoch 7/100
175/175 [==============================] - 1s 3ms/step - loss: 0.1518 - accuracy: 0.9349 - val_loss: 0.3161 - val_accuracy: 0.8968
```

Fig 8.9 Running of epochs after input layering and model fit

turnitin

Chirag Verma, Devansh Agarwaal MAJOR PROJECT REPORT.pdf
May 31, 2021
10510 words / 55536 characters

Chirag Verma, Devansh Agarwaal MAJOR PROJECT REPORT.pdf

Sources Overview

13%

OVERALL SIMILARITY

1	Jacobs University, Bremen on 2019-08-20 SUBMITTED WORKS	3%
2	pragrammesatvqms.blogspot.com INTERNET	1%
3	University of the Pacific on 2020-02-01 SUBMITTED WORKS	<1%
4	Middlesex University on 2020-04-15 SUBMITTED WORKS	<1%
5	wildy.com INTERNET	<1%
6	www.scribd.com INTERNET	<1%
7	Intercollege on 2018-11-06 SUBMITTED WORKS	<1%
8	www.investopedia.com INTERNET	<1%
9	RMIT University on 2019-06-06 SUBMITTED WORKS	<1%
10	blog.atlasrfidstore.com INTERNET	<1%
11	RMIT University on 2020-02-02 SUBMITTED WORKS	<1%
12	Higher Education Commission Pakistan on 2012-06-30 SUBMITTED WORKS	<1%
13	nguyendang.net.vn INTERNET	<1%
14	www.practicalpointers.org INTERNET	<1%
15	Higher Education Commission Pakistan on 2020-04-21 SUBMITTED WORKS	<1%
16	Parikshit N. Mahalle, Prashant S. Dhotre. "Chapter 5 Architecture for Context-Aware Systems", Springer Science and Business Media LL... CROSSREF	<1%
17	University of Greenwich on 2021-05-11 SUBMITTED WORKS	<1%

18	scholarsmine.mst.edu INTERNET	<1%
19	www.gtu.ac.in INTERNET	<1%
20	Thiagarajar School of Management on 2016-08-09 SUBMITTED WORKS	<1%
21	madoc.bib.uni-mannheim.de INTERNET	<1%
22	www.rdocumentation.org INTERNET	<1%
23	Universiteit van Amsterdam on 2018-07-06 SUBMITTED WORKS	<1%
24	erepo.usiu.ac.ke INTERNET	<1%
25	repository.its.ac.id INTERNET	<1%
26	serokell.io INTERNET	<1%
27	www.researchgate.net INTERNET	<1%
28	City University on 2019-09-17 SUBMITTED WORKS	<1%
29	University of Teesside on 2021-05-11 SUBMITTED WORKS	<1%
30	biomedpharmajournal.org INTERNET	<1%
31	docshare.tips INTERNET	<1%
32	medinform.jmir.org INTERNET	<1%
33	repositorio.furg.br INTERNET	<1%
34	repositorio.uchile.cl INTERNET	<1%
35	www.mdpi.com INTERNET	<1%
36	www.savap.org.pk INTERNET	<1%
37	CSU, Fullerton on 2020-04-24 SUBMITTED WORKS	<1%
38	Sheffield Hallam University on 2019-09-03 SUBMITTED WORKS	<1%
39	Tilburg University on 2021-05-20 SUBMITTED WORKS	<1%
40	Trinity Grammar School on 2020-08-10 SUBMITTED WORKS	<1%
41	download.atlantis-press.com INTERNET	<1%

42	www.iaescore.com INTERNET	<1%
43	Duy Nguyen Duc, Narameth Nananukul. "A Hybrid Methodology Based on Machine Learning for a Supply Chain Optimization Problem", ... CROSSREF	<1%
44	Feras A. Batarseh, Gayatri Nambiar, Gerald Gendron, Ruixin Yang. "Geo-Enabled Text Analytics through Sentiment Scoring and Hierarch... CROSSREF	<1%
45	University of Melbourne on 2020-06-06 SUBMITTED WORKS	<1%
46	insis.vse.cz INTERNET	<1%
47	www.hindawi.com INTERNET	<1%

Excluded search repositories:

- None

Excluded from Similarity Report:

- Bibliography
- Quotes
- Small Matches (less than 10 words).

Excluded sources:

- None