# Application of Instance Learning Algorithms to Analyze Logistics Data

Urenna Nwagwu
Wichita State University
Kansas City, USA

Abiodun Ayinde
Community Health Network
Indianapolis, USA

Yemisi Olasoji
Quadrilateral Technologies Limited
Ibadan, Nigeria

***Abstract:* - Predictive analytics has been applied in all works of life with the aim to process, analyze, evaluate, mitigate risks and create different computational models based on the organization goals for adequate decision making when a proper knowledge discovery methodology is adopted by the predictive learning environment. Several instance based learning classifiers have been applied in the field of supply chain management to build a multilayer learning predictive platform but with little emphasis on classifiers function normalization to reduce the convergence time of the trained model. A lot of companies have benefited by automating predictive analytics for their workflow, resource sharing, decision making, etc. With the evolution of various cloud computing service, startups companies have tapped into the cloud service to scale their business by leaning on enterprise based predictive models to drive their workflow for vital decision making process at lower cost to improve their forecasting abilities and responsiveness via real-time analytics.**

***Keyword: - Logistic, Instance Based Learning, Modeling***

## 1. INTRODUCTION

Predictive analysis is used to make forecasts by reading algorithms based on both current and historical data. Organizations can modify how and where they use resources to better prepare for future events. It creates a framework for connecting the dots between trends, patterns, and associations in data to help businesses respond proactively to future developments.

Predictive analytics uses a mixture of analytics methodologies, combined with automated tools and technologies, to find patterns within data, and goes one step further to anticipate the likelihood of specific future events.

Supply chain management creates values through changes in time, location, quantity and quality and has the potential for huge competitive advantage for the Organization. Major elements of the supply chain management have always tended to reflect supplier, storage, manufacturing, distributor, retailer and customer without the other important elements like logistics, traffic management and quality control amongst others. Often, poor logistics arrangements have posed serious problems for effective supply chain management in Nigeria.

## 2. LITERATURE REVIEW

A multi-stage hybrid model for analyzing a supply chain network (SCN) collapse recovery possibility, the first stage of the model, analyzed the ability of an SCN to fulfil its customers required due date (RDD). As soon as the SCN defaults to timely fulfil their customers RDD, the second stage of the model is triggered to measure the collapse recovery possibility (CRP) of the SCN. Final stage calculate the collapse recovery possibility (FCRP) of the SCN. Since the operation times, customer demand and external supply of raw material are uncertain, fuzzy triangular numbers to estimate the value of foregoing parameters. Subsequently, fuzzy program was employed to evaluate and review technique (FPERT) to calculate the completion time of SCN operations. In the third stage, for the critical elements of the SCN obtained from FPERT, the SC simulator is developed to provide a dynamic view of the SC and assesses the impact of decisions recommended by the SC fuzzy models on SC performance. Moreover, an empirical example is presented to validate the effectiveness of the proposed model. Finally, sensitivity analysis was conducted on the parameters employed in the model to analyze the behavior of each parameter [1]

The volume of data generated from all parts of the supply chain has changed the nature of SCM analysis. By increasing the volume of data, the efficiency and effectiveness of the traditional methods have decreased. Limitations of these methods in analyzing and interpreting a large amount of data have led scholars to generate some methods that have high capability to analyze and interpret big data. Therefore, the main purpose of this paper is to identify the applications of machine learning (ML) in SCM as one of the most well-known artificial intelligence (AI) techniques. By developing a conceptual framework, this paper identifies the contributions of ML techniques in selecting and segmenting suppliers, predicting supply chain risks, and estimating demand and sales, production, inventory management, transportation and distribution, sustainable development (SD), and circular economy (CE) [2]

Application of model predictive control (MPC), an advanced control technique originating from the process industries, to supply chain management (SCM) problems arising in semiconductor manufacturing. This work is to demonstrate the usefulness of MPC as a tactical decision policy that is an integral part of a comprehensive hierarchical decision framework aimed at achieving operational excellence. A fluid analogy is used to describe the dynamics of the supply chain. Compared to traditional flow control problems, challenges of SCM in semiconductor manufacturing result from high stochasticity and nonlinearity in throughput times, yields and customer demands. The advantages of the control-oriented receding horizon formulation behind MPC are presented for three benchmark problems which highlight distinguishing features of semiconductor manufacturing. The effects of tuning, model parameters, and capacity are shown by comparing system robustness and multiple performance metrics in each case study [3]

Value chain management has gone through various stages of automation, integration and optimization in the past decades. While an optimization model for value chain deals with business scenarios under known circumstances, a predictive value chain model deals with probable circumstances in the future. Predictive analytics is succeeding optimization in the evolution of technologies supporting value chain management. This paper proposes a forward looking value creation model that combines the important concepts of value chain management and predictive analytics. An enterprise model for value chain predictive analytics that facilitates the convergence of information, operations and analytics is presented [4]

The comparative analysis of the three predictive classifiers showed that IBk gave the best Precision Accuracy with the lowest null hypothesis value. Also, True Positive Rate (TPR) of Naïve Bayes and Classification And Regression Tree decreases with decreasing value of the Precision Accuracy while TPR of IBk increases with increasing value of PA, which implies that IBk performs best with increasing value of True Positive Rate [5]

## 3.0 METHODOLOGY

### *3.1 Overview*

Predictive analysis scope in big data analytics plays an important role in forecasting and reveal some insight in both background study and model experimentation of adopted data mining techniques. This study adopted and compared the performance evaluation of two instance based learning algorithms namely; k-nearest neighbor (KNN) and radial basis network function (RBN) to logistics data extracted from Koloexpress logistics database. Rapid miner was adopted as the machine learning platform for this study. The predictive model was built to predict the average delivery time of an end to end delivery within Ibadan metropolis and the performance of the algorithms were evaluated. The learning architecture of the model was calibrated for both algorithms at 0.01 intervals and the time prediction was factored into the model to track the divergence time by averaging the actual time and expected delivery time.

### *3.2 Data*

The logistics data were obtained from Koloexpress database from its Ibadan office. Koloexpress is an e-commerce and logistics company situated in Ibadan, Oyo State, Nigeria. One of their services encompassed an end to end delivery service across the world. The data for this research was streamlined to an end to end deliveries within Ibadan metropolis.

### *3.3 Research Approach*

The Cross-Industry Standard Process for Data Mining (CRISP- DM) model was implemented as a standard approach for this research work. It is a cyclic approach that consists of six main phases– Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment. The total of 3,000 dataset served as the input data at the modeling stage of the CRISP - DM. The data was partitioned in training and texting data, comprising of 80percent and 20percent respectively. As shown in figure 1

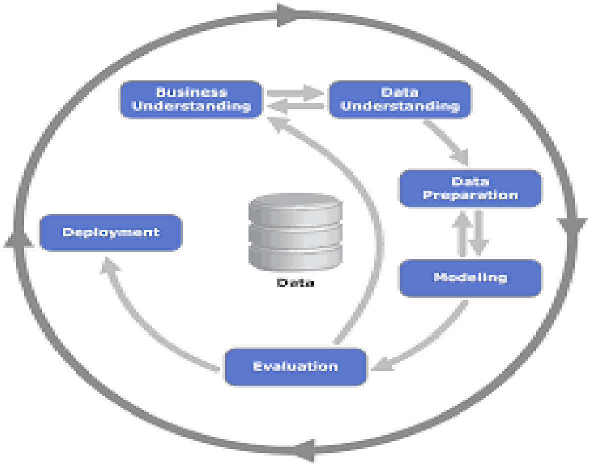


*Figure 1: Cross Industry Standard Process for Data Mining*

### *3.4 Equations*

KNN: applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function f (x) can take on any value from some finite set V. A set of training examples of tuple of attribute value ($a_1$, $a_2$ …. $a_n$).The learner is asked to predict the target value or classification for this new instance. The nearest neighbor approach shall be used to classify the new instance to be assigned the most probable target, $V_{map}$, given the attribute value ($a_1$, $a_2$……$a_n$ that describe the instance.

$V_{map} = \arg_{max} P(V_j \quad a_1, a_2, \ldots\ldots a_n)$ …………….. (1)
$V_{map} = P(V_j | a_1, a2, \ldots\ldots a_n | V_j)\, P(V_j)$ ……………... (2)

RBN: Radial basis network function (RBN) typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer. The input can be modeled as a vector of real numbers. The output of the network is then a scalar function of the input vector. Below are the final derived formulae for the radial basis network.

$x(t+1) = f[x(t)] = 4x(t)\,[1-x(t)]$ ………………………… (3)

$x(t+1) = f[x(t)] = \text{ɥ}(t) = \text{ɥ}[x(t)]$ ………………………… (4)

$\sum = f[x(t)] - \text{ɥ}[x(t)] = x(t+1) - c[x(t), t] - \text{ɥ}[x(t) = x(t+1) - d(t+1)$ ……………………………………………………… (5)

c = actual pickup time
d = expected delivery time
t = average delivery time
ɥ = learning value

## 4.0 ANALYSIS OF RESULT

Table1: Learning Value = 0.01

| METRICS | KNN | RBN |
|---|---|---|
| True Positive Rate | 0.47 | 0.67 |
| False Positive Rate | 0.38 | 0.20 |
| Precision Rate | 0.77 | 0.68 |

Table2: Learning Value = 0.02

| METRICS | KNN | RBN |
|---|---|---|
| True Positive Rate | 0.86 | 0.84 |
| False Positive Rate | 0.46 | 0.10 |
| Precision Rate | 0.87 | 0.92 |

Table3: Learning Value = 0.03

| METRICS | KNN | RBN |
|---|---|---|
| True Positive Rate | 0.55 | 0.69 |
| False Positive Rate | 0.20 | 0.45 |
| Precision Rate | 0.91 | 0.82 |

Table4: Learning Value = 0.04

| METRICS | KNN | RBN |
|---|---|---|
| True Positive Rate | 0.37 | 0.51 |
| False Positive Rate | 0.18 | 0.20 |
| Precision Rate | 0.89 | 0.96 |

## 5.0 RESULT DISCUSSION

Instant learning classifiers such as: K-Nearest Neighbor (KNN) and Radial Basis Network (RBN) were calibrated on Rapid Studio Block in Rapid Miners to train and test the developed model. A dataset of 3000 logistics data were created, each with 10 attributes. The attributes were sex, age, local government area (LGA), state, pickup time, delivery time, item weight, item category, delivery status and mode of payment. Nine (9) of the attributes served as the input (sex, age, local government area (LGA), state, pickup time, item weight, item category, delivery status and mode of payment.) in predicting the output attribute (delivery time).These classifiers were calibrated using instant memory learning methodology (80percent of the dataset for training and 20percent for testing) using the rapid miner studio for preparing, modeling for the result analysis The performance of the classifiers was evaluated based on: True Positive Rate (TPR), False Positive Rate (FPR) and Performance Accuracy (PA). The Tables revealed that the RBN outperformed KNN when the learning value was set at 0.02. RBN has the highest TPR and PR with the lowest FPR at 0.02 because RBN predicted the next instance of the class by making predictions just-in-time by calculating the similarity between an input sample and each training instance for every value of the target variable (delivery time). The highest precision rate of RBN was based on the fact that it took milliseconds for the model to process, learn and converge compare to KNN which took approximately three minutes for the model to learn and converge when predicting the delivery time.

Both classifiers performed optimally when iterative and incremental learning were adopted in predicting the target variable but KNN under performed when calibrated for instance memory learning compared to RBN. Future works, should look into hybridizing a neural network with KNN to reduce the learning and convergence period of the model, thereby improving the performance of instance memory learning predictions for some classifiers.

## REFERENCES

[1] B. Vahdani, M. Zandieh and V. Roshanei, "A hybrid multi-stage predictive model for supply chain network collapse recovery analysis: a practical framework for effective supply chain network continuity management" International Journal of Production Research, Volume 49- 2011, Issue 7, pp 2035 – 2060, April 2010

[2] W. Wang, D.E Rivera, K.G Kemp, " Model predictive control strategies for supply chain management in semiconductor manufacturing" International Journal for Production Economics, Volume 107, Isuue 1, pg 56 – 77, May 2007

[3] S. Sadeghi, F. Mansoon, H. Rezaei, S. Aeini, "Application of Machine Learning in Supply Chain Management: A comprehensive overview of the main areas" Mathematical Problems in Engineering / 2021 / Article, June, 2021

[4] J. O Chan " A predictive analytic model for value chain management" Journal of International Technology and Information, Volume 16, Issue 1, 2007

[5] A.Q Ayinde, A.B Adetunji, E.O Omidiora " Comparative analysis of some selected classifiers in mining students' educational data", Communications on Applied Electronics (CAE) – ISSN : 2394-4714 Foundation of Computer Science FCS, New York, USA Volume 1 – No.5, April 2015 – www.caeaccess.org