SUST Journal of Engineering and Computer Sciences (JECS), Vol. 20, No. 3, 2019

A Type-2 Fuzzy Logic Based System for Decision Support to Minimize Financial Default in the Sudanese Banking Sector

Ahmed Salih [1] and Hani Hagras[2]

[1] Department Computer Science, College of Postgraduate Studies, Sudan University of Science and Technology (SUST), salihazez@yahoo.com

[2]The Computational Intelligence Centre, School of Computer Science and Electronic Engineering University of Essex, United Kingdom, hani@essex.ac.uk

Received:20/09/2019
Accepted: 15/12/2019

ABSTRACT—The recent global financial-economic crisis has led to the collapse of several companies from all over the world. This has created the need for powerful frameworks which can predict and reduce the potential risks in financial applications. Such frameworks help organizations to enhance their services quality and productivity as well as reducing the financial risk. The widely used techniques to build predictive models in the financial sector are based on statistical regression, which is deployed in many financial applications such as risk forecasting, customers' loan default and fraud detection. However, in the last few years, the use of Artificial Intelligence (AI) techniques has increased in many financial institutions because they can provide powerful predictive models. However, the vast majority of the existing AI techniques employ black box models like Support Vector Machine (SVMs) and Neural Network (NNs) which are not able to give clear and transparent reasoning to explain the extracted decision. However, nowadays transparent reasoning models are highly needed for financial applications. This paper presents a type-2 fuzzy logic system for predicting default in financial systems. the researchers used a real dataset collected from the banking sector in Sudan. The proposed system resulted in transparent outputs which could be easily understood, analyzed and augmented by the human stakeholders. Besides, the proposed system resulted in an average recall of 83.5%, which outperformed its type-1 counterpart by 20.66%.

Keywords**:** *Type-2 fuzzy logic system, default, prediction model.*

المستخلص– إن الأزمة الإقتصادية العالمية قد أرثت آثار سالبة على معظم الشركات العالمية، ومن هنا أتت الحوجة الملحة إلى أنظمة للتنبؤ بالمخاطر بغرض التقليل من حدة آثارها على الاقتصاد العالمي. يمكن لهذه الأنظمة أن تسهم في تحسين جودة وانتاجية الخدمات المقدمة بواسطة المؤسسات المالية، وفي نفس الوقت يمكن أن تسهم في تقليل حدوث المخاطر. تعد تقنية الإنحدار الإحصائي من أكثر التقنيات المستخدمة في بناء أنظمة التنبؤ . في الآونة الأخيرة بدأ استخدام تقنيات الذكاء الاصطناعي في مجال التنبؤ بالأزدياد لأن الأنظمة المبنية بواسطة هذه التقنيات أثبتت قوتها. ولكن معظم التقنيات التي تم استخدامها مع قوتها إلا أنها تعتبر من نوع الصندوق الأسود والتي لا توضح أسباب التبؤ . في الأونة ألاخير أصبح توضيح الأسباب التي أدت إلى التبؤ من الأشياء المهمة والتي من شأنها أن تخدم متخذي القرار وذلك بمدهم بمعلومات من شأنها أن تكشف الأسباب التي تؤدي إلى مثل هذه المخاطر.في هذا البحث تم استخدام تقنية المنطق الضبابي النوع الثاني (type–2 fuzzy logic) لبناء نظام التنبؤ بخطر التعثر في الأنظمة المصرفية. تم استخدام قاعدة بيانات حقيقية من بنك الشمال السلامي السوداني . النظام المقترح بعد تطبيقه انتج نموذج يعتبر من نماذج الصندوق الأبيض والتي يمكن أن توضح وتكشف أسباب التعثر. أعطى النظام المقترح نسبة دقة في التبوء 83,5 وبذلك يتفوق على نموذج المنطق الضباب النوع الأول (type–1 fuzzy logic) بنسبة 20,66 %.

I. *INTRODUCTION*

During 2008 economy crisis, several companies financially collapsed around the world. For example, the United State housing market lost \$3.4 Trillion in real estate wealth [1]. This was equivalent to \$30,300 per U.S. household. Stock wealth lost $7.4Trilion equivalent to $66,200 per household. 5.5 Million jobs were lost in the American job market. All of these factors have taken hold despite the existent of predictive models to help forcast crisis before they happen.

Figure 1 shows the impact of the crisis on wages in the US between 2007 and 2009. More emphasis on finding ways to minimize the impact of potential risks on businesses becomes evident among reasearchers. An example of technique being adopted to accurately predict risks and impact is the use of the Artificial Intelligence (AI) [2].

Advancement in online technology have enabled organizations to access massive amount of data to perfect prediction models. Recent hardware technologies have made it cheaper to store and analyze these vast amounts of data and in short time spans. Therefore many financial organizations attempted to build accurate predictive models to mine available historical data and extract relevant indicators and produce better decisions on financial operations[3].

Traditional statistical models have been used in the financial sector for long time including using logistic regression to predict banks failure and firms failure [4-5]. Banks et. al. [6] have developed simple linear models to classify loan risks and predict commercial bank failure in Turkey. West et. al. [7] found that the factor analysis and logit estimation combination is a promising method for evaluating bank condition. Neural Networks (NNs) were also used widely for bankruptcy prediction and was compared with DA, factor logistic, K-NN and ID3. It has been shown that the (NNs) perform better than other techniques in terms of predictive accuracy [8-11].

Statistical techniques are widely used because they are easy to develop. However, they assume a certain mathematical relationship between the input and the output, which is not the case for the majority of real world data [2].

In this paper an overview of predictive models for financial applications, development and implementation of type-2 fuzzy logic system, will be presented. The model is evaluated and validated with real time data extracted from Sudanese banking sector. This is an unprecedented work that exercise real finance data from Sudan banking sector. The Sudanese banking sector lacks use of predictive models for decision support, and hence have suffered from defaults. The paper will conclude with a list of findings and recommendations.

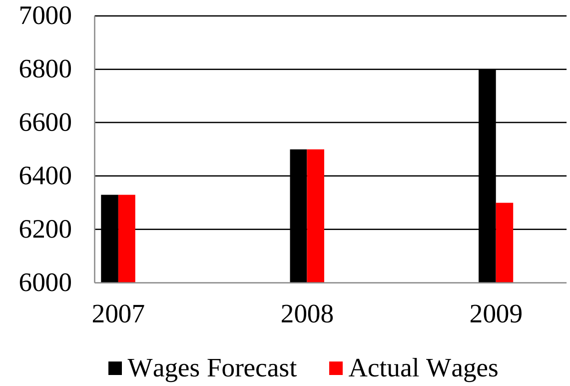

Figure 1. Impact of the economic crisis on wages in US with existence prediction techniques [1]

II. Predictive Models for Financial Applications Overview

In general, there are four different techniques to build predictive models employed by financial firms. There are:

- Statistical-based
- Operation research-based
- Artificial Intelligence (AI) based
- Hybrid artificial intelligent based.

The statistical-based predictive models contain many techniques like:

- Dicremenant Analysis (DA),
- Statistical Regression (SR)
- Factor Analysis (FA)

The techniques are wildly used because they are easy to develop. However, they capture only information that can be used within mathematical models. The output in this case is binary either a 0/1 or black/white [2]. Moreover, these techniques assume existence of mathematical relationship between input and output which is necessarily true in the real-world data.

Operation research-based predictive models contain many techniques like:

- Linear Programing(LP),
- Data Envelopment Analysis(DEA)
- Quadratic Programing (QP)

These techniques are used widely because of their development simplicity but they are complicated to use and can lead to complicated semi-black box models.

Artificial Intelligence (AI)-based predictive models can be subdivided into two sections:

- **Black-box:** containing techniques such as Support Vector Machine (SVM) [7] and Neural network (NN) [8]
- **White-box:** containing techniques such as Case Base Reasoning (CBR), Rough Set Theory, Decision Tree (DT), and Fuzzy logic (FL) [3].

The **Black Box** models are used on a wide spectrum of financial applications such as [12] and they produce a good level of prediction accuracy. However, these models are hard to understand and analyze by financial analysts since black-box models don not produce clear evidence-based decisions. This is considerd an important requirement by the financial market nowadays due to the intense competition and race to winning customer confidence.

The term **White Box** refers to AI techniques that can provide transparent reasoning behind extracted decisions. This has motivated users to apply white box techniques for normal end usages [3]. The White box models uses the following techniques:

- Case-based Reasoning
- Decision Trees
- Fuzzy Logic

Figure 2 shows a schematic of white-box model where decisions are provided with justifications. In the following section a summary of these techniques will be provided. The Case-Based Reasoning (CBR) is one of the white box methods that attempts to solve new problems based on solutions of similar past problems [13-14].

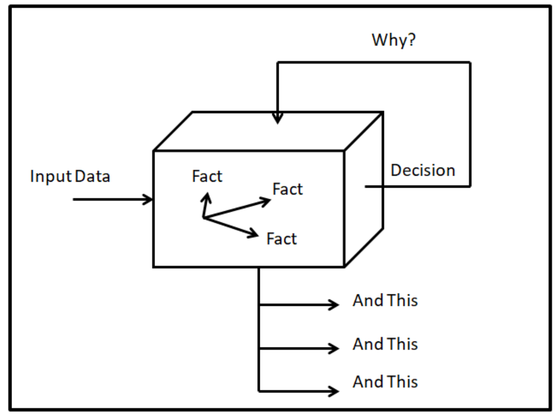

Figure 2. A simple visualization for white box model

Decision trees (DT) uses recursive partitioning algorithm to produce rules on a specific data set [15]. DT algorithms take training data set and extract decision boundaries. These decision boundaries are then used to build a decision tree. From the constructed decision tree, the model will be able to extract decision rules which can provide reasoning tools which can provide clear understanding about the extracted decisions. However, DTs have many limitations such as inability to handle uncertainty. DT also utilize recursive partitioning operation which may leads to hard decision boundary extraction [16].

Lotfi Zadeh [17] have proposed the Fuzzy Logic Theory (FL) in order to provide a framework that is capable of handling uncertainties associated with natural languages. The FL tries to mimic a human's way of reasoning in order to think in approximate ways rather than precise ways. The FL systems is built based on fuzzy set theory which provides means of calculating intermediate values between absolute true and absolute false. The resulting values range between 0 and 1 leading to smooth transition between different sets. Fuzzy Logic Systems (FLSs) have been employed widely in financial applications.

J. Andres et. al. [18] have proposed fuzzy-rule-based classifiers for bankruptcy prediction problem and compared their classifier with logit and perceptron NN techniques. It was concluded that NN and fuzzy rule-based classifier outperformed logistic regression. The FL can provide transparent reasoning model; however, type-1 FL cannot handle a high level of uncertainty. The FL systems suffer from the dimensionality problem where number of rules tends to be enormous. This render them tedious to read and analyze by humans.

The term (Hybrid Intelligent Technique) refers to the AI technique which tries to combine more than one AI technique to take advantage of each technique individual features and overcome each system limitation. S. Michael et. al.[19] presented the combined use of a fuzzy rule generation method and a data mining technique for the assessment of financial risks. A comparison between developed model with DA, logit analysis, and probability analysis concluded that fuzzy rule-based classifier outperformed other methods.

III. Type-2 Fuzzy Sets and Systems

Type-2 Fuzzy sets initially introduced by L. Zadeh[17] in 1975 as an extension of Type-1 fuzzy set. The membership grades of the Type-2 fuzzy sets are of Type-1 fuzzy sets. These Type-2 fuzzy sets are very useful when it is difficult to determine an exact membership function as in Type-1 fuzzy sets [20]. When there is no membership uncertainity, the set reduces to Type-1 fuzzy set.

Figure 3 shows Type-2 fuzzy set which is characterized by a fuzzy Membership Function (MF). The MF will assume membership value (or membership grade) for each element on the fuzzy set between [0, 1]. The grade will be of an interval set of values rather than a single value. In contrast, the Type-1fuzzy set the membership grade is a crisp and single value falling between “0” and “1”.

From Figure 3, when the Upper Membership Function and Lower Membership Function coincides, the Figure reduces to Type-1 Fuzzy set. This indicates that the FOU region is eliminated in Type-1 Fuzzy set in accordance to the definition.

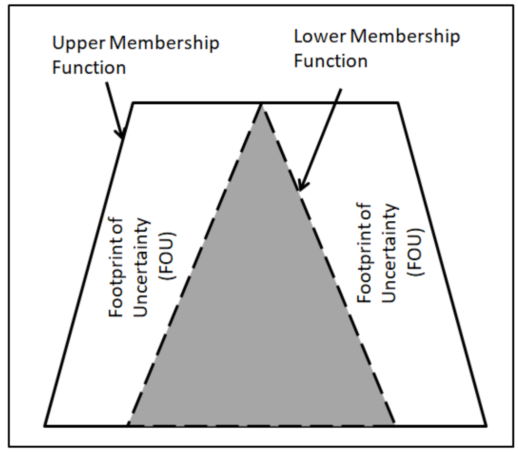

Figure 3: A Type-2 Fuzzy Set

A Type-2 fuzzy set is bounded from bottom by a Lower Membership Function and bounded from top by an Upper Membership Function. The membership functions of Type-2 fuzzy sets are 3D dimensional and include a Footprint Of Uncertainty (FOU). The combination of the Type-2 third-dimension and the FOU provides additional degrees of freedom that enables direct modelling and uncertainties handling.

In Type-2 fuzzy logic system each input and output will be represented by a large number of Type-1 fuzzy sets, which are embedded in the Type-2 fuzzy sets. The concept of a principal membership function also illustrates the fact that a Type-1 fuzzy set can be thought of as a special case of a Type-2 fuzzy set. We can think of a Type-1 fuzzy set as a Type-2 fuzzy set whose membership grades are Type-1 fuzzy singletons. Also, having secondary membership equal to unity for only one primary membership and zero for all others [21].

In Figure 4, the structure of a standard Type-2 Fuzzy Logic System (FLS) is presented. The crisp inputs are first fuzzified i.e. inputs are converted to input Type-2 fuzzy sets. Then, the inference engine identifies the rules fired from a previously defined rule base. Then combining these rules to produce output Type-2 fuzzy sets. Subsequently, the Type-2 fuzzy output sets are reduced and mapped to Type-1 fuzzy sets.

This process is also known as type-reduction technique indicated by the Type Reducer block in Figure 4. In this process, the Type-2 fuzzy sets outputs are reduced to Type-1 fuzzy sets by performing centroid calculation. Finally, the Type-1 reduced fuzzy sets are defuzzified i.e. by taking the average of the type-reduced set to obtain a crisp output [22-23].

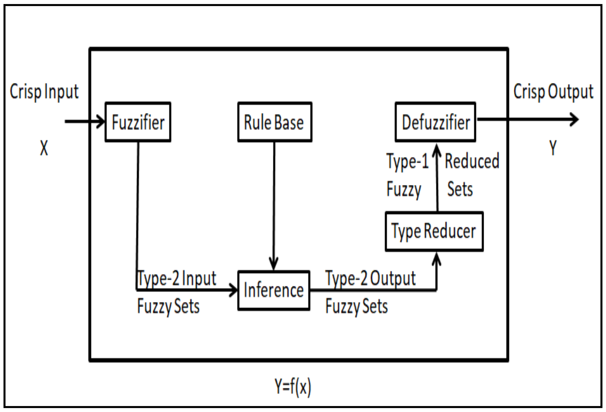

Figure 4: Standard Type-2 FLS

Type-1 FLSs cannot fully handle or accommodate the high levels of linguistic and numerical uncertainties. This is due to usage of precise Type-1 fuzzy sets and membership functions. For example, for a house environment, a "Moderate" temperature could be associated with the triangular Type-1 fuzzy membership function.

However, the center of this triangular membership function and its endpoints vary according the user of the system where different users will have different preferences. Even for the same user, his/her preference will vary according to the season of year, his mode, country, context, and room location in the house. For example, "Moderate" temperature in the kitchen will be different to "Moderate" temperature in the living room.

IV. Proposed Type-2 Fuzzy Logic-Based System

The proposed model is an implimintation of system this system take the customer's information as an input and provides classification of this customer(default/ not default). Figure 5 shows the structure for the proposed type-2 fuzzy logic based system for decision support to minimize financial defaults in the Sudanese banking sector. The proposed model is divided in two phases, the modeling phase and the prediction phase. The details of these phases will be discussed in detail in the following subsections.

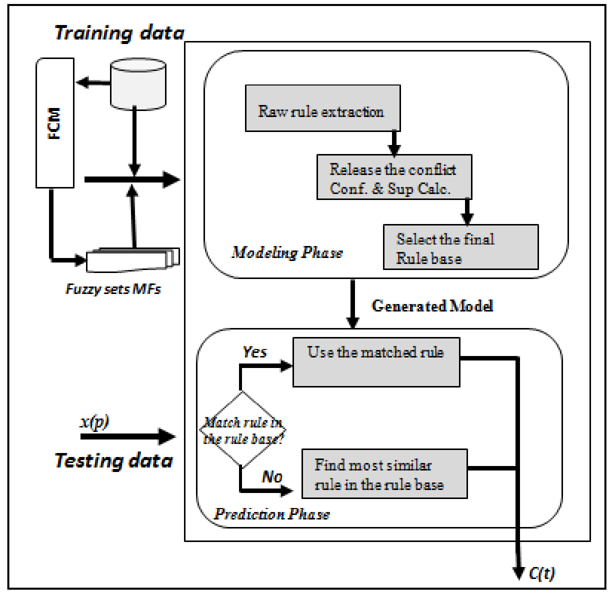

Figure 5: Flow diagram of proposed model

(A) Modeling phase:

In this phase there are two components which must be constructed. These are the fuzzy sets Membership Function (MFs) and the rule base. In order to build the MFs for each continuous input parameter in the data set we used the fuzzy C-Means clustering algorithm (FCM) [24-25]. Fuzzy c-means (FCM) is a clustering method that allows each data point to belong to multiple clusters with varying degrees of membership. FCM is based on the minimization of the following objective function [24]:

$$j_m = \sum_{i=1}^{D} \sum_{j=1}^{N} \mu_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad (1)$$

where D is the number of data points, N is the number of clusters, m is fuzzy partition matrix exponent for controlling the degree of fuzzy overlap, with m>1. Fuzzy overlap - which refers to how fuzzy the boundaries between clusters are - that is the number of data points that have significant membership in more than one cluster. The xi is the ith data point and cj is the center of the jth cluster. μij is the degree of membership of xi in the jth cluster. For a given data point, xi, the sum of the membership values for all clusters is one.

The FCM operates as follows:

1. Randomly initialize the cluster membership values, μij.

2. Calculate the cluster centers [25]:

$$c_j = \frac{\sum_{i=1}^{D} \mu_{ij}^{m} x_i}{\sum_{i=1}^{D} \mu_{ij}^{m}} \quad (2)$$

3. Update μij according to the following[23]:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{N} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (3)$$

4. Calculate the objective function, Jm.
5. Repeat steps 2–4 until Jm improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

To generate the rule base, we use the following steps:

Step1: Raw Rule Extraction:

From the training data set which contains numbers of record each one represents input-output pair (x(t), C(t)), t =1,...,T (T is the total number of training dataset records available for the training phase) as shown in Figure 6.

- Calculate the upper and lower membership values ($\bar{\mu}_{A_s^q}$, $\underline{\mu}_{A_s^q}$) for any antecedent fuzzy set q=1,…K (*K* is the total number of fuzzy sets representing the input pattern s where *s=1…n.*).
- Generate all possible rules that can be extracted from each input-output pair (x(t), C(t)) as result number of rules will generated form the single input-output pair (x(t), C(t)) with same consequence C(t) but different antecedences each of them could be written as follows:

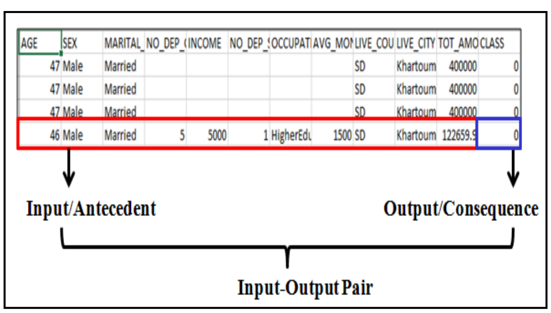

Figure 6. Representation of input-output data pairs

$$R^j: If\ x_1 is\ \tilde{A}_1^{qjt}\ and \dots and\ x_n\ is\ \tilde{A}_n^{qjt}$$
$$then\ class\ C_t, t = 1,2 \dots T \quad (4)$$

- To measure the strength of the point x(t) belonging to the fuzzy area that covered by the generated rule we need to calculate the firing strength Ft which is defined with its lower and upper bound ($\overline{f^{(t)}}$, $\underline{f^{(t)}}$) These bounds can be calculated as follows:

$$\underline{f}^{jt}(x^{(t)}) = \underline{\mu}_{A_1^{qjt}}(x_1) \dots .\ \underline{\mu}_{A_n^{qjt}} \quad (5)$$

and

$$\overline{f}^{jt}(x^{(t)}) = \overline{\mu}_{A_1^{qjt}}(x_1) * \dots .* \overline{\mu}_{A_n^{qjt}}(x_n) \quad (6)$$

The (*) represents the minimum or product t-norm.

The resarchers repeat Step1 for all input-output pairs in training data set and this resulted in a number of rules. Each will be extracted using the form of Equation (4).

Step 2: Resolve The Conflict by Using Confidence and Support:

Some of extracted rules have a same antecedent and different consequences which can causes conflicts for the fuzzy logic system in the prediction phase. For example, consider we have two rules like:

- *R1: if x1 is low and x2 is mid then class is* ***good***
- *R2: if x1 is low and x2 is mid then class is* ***bad***

The problem here will be which one of these rules we will need to be used to predict customer status. It is evident that the customer cannot be good and bad as the same time. To resolve this conflict, one would need to replace any group of rules which share the same antecedent with one single rule.

Another issue that needs to be considered is the competition fairness between the two consequences classes. This is due to the fact that financial data is imbalanced by nature which means the majority class is good customer and the minority class is bad customer. Thus in order to resolve the rule's conflict, we use the "***weighted scaled dominance***" approach introduced by [2] and "***weighted confidence***" which is presented by [26] using the following steps:

a) To calculate the scaled dominance for a given rule that had the consequence class C_j, we divide the firing strength of this rule by the summation of the firing strengths of all rules which had consequent class C_j. We scale the firing strength by scaling its upper and lower bounds as follows:

$$\overline{fs^{Jt}} = \frac{\overline{f^{Jt}}}{\sum_{J\epsilon\, class} \overline{f^{Jt}}} \quad (7)$$

$$\underline{fs^{jt}} = \frac{\underline{f^{jt}}}{\sum_{J\epsilon\, class} \underline{f^{jt}}} \quad (8)$$

b) To resolve the conflict in the consequence for rules that share the same antecedence we need to calculate the **scaled confidence** and **scaled support**. The **scaled confidence** ($\tilde{A}_q \rightarrow C_q$) for given "**m**" rules having the same antecedents and conflicting classes is defined by its upper bound $\bar{c}$ and lower bound $\underline{c}$ could be written as follows [26]:

$$\bar{c}(\tilde{A}_q \rightarrow C_q) = \frac{\Sigma_{xs\,\epsilon\, class\, Cq}\overline{f^{jt}x(s)}}{\sum_{j=1}^{m}\overline{f^{jt}x(s)}} \tag{9}$$

$$\underline{c}(\tilde{A}_q \rightarrow C_q) = \frac{\Sigma_{xs\,\epsilon\, class\, Cq}\underline{f^{jt}x(s)}}{\sum_{j=1}^{m}\underline{f^{jt}x(s)}} \tag{10}$$

The ***scaled support*** is defined by its upper bound $\bar{s}$ and lower bound $\underline{s}$ can be calculated as follows [25]:

$$\bar{s}(\tilde{A}_q \rightarrow C_q) = \frac{\Sigma_{xs\,\epsilon\, class\, Cq}\overline{f^{jt}x(s)}}{m} \tag{11}$$

$$\underline{s}(\tilde{A}_q \rightarrow C_q) = \frac{\Sigma_{xs\,\epsilon\, class\, Cq}\underline{f^{jt}x(s)}}{m} \tag{12}$$

c) The scaled dominance (which is defined by its upper bound $\bar{d}$ and lower bound $\underline{d}$) can be calculated now by multiplying scaled support and scaled confidence of the rule as follows [2]:

$$\bar{d}(\tilde{A}_q \rightarrow C_q) = \bar{s}(\tilde{A}_q \rightarrow C_q) * \bar{c}(\tilde{A}_q \rightarrow C_q) \tag{13}$$

$$\underline{d}(\tilde{A}_q \rightarrow C_q) = \underline{s}(\tilde{A}_q \rightarrow C_q) * \underline{c}(\tilde{A}_q \rightarrow C_q) \tag{14}$$

d) Then we need to calculate the **"weighted scaled dominance"** (which is defined by its upper bound $\overline{wd}$ and lower bound $\underline{wd}$) as follows [2]:

$$\overline{wd}(\tilde{A}_q \rightarrow C_q) = \bar{d}(\tilde{A}_q \rightarrow C_q) * \overline{d_{ave}} \tag{15}$$

$$\underline{wd}(\tilde{A}_q \rightarrow C_q) = \underline{d}(\tilde{A}_q \rightarrow C_q) * \underline{d_{ave}} \tag{16}$$

where d_{ave} is the average dominance (defined in terms of $\overline{d_{ave}}$ and $\underline{d_{ave}}$) over fuzzy rules with the same antecedent $\tilde{A}_q$ but different consequent classes C_q.

e) Finally replace the group of rules which share the same antecedence but different class with one rule that have antecedence that shared by this group and the consequent class which will be corresponding to the rule that gives the highest average weighted scaled dominance

$$\textit{\textbf{Value}} = \frac{\overline{wd} + \underline{wd}}{2}$$

(B) Prediction Phase:

So far, we have a full Type-2 fuzzy logic classifier which is ready to accept new input patterns with unknown classes and predict which classes these input belong to. When a new input pattern $x^{(p)}$ is presented to the proposed model, then there is one of two possible cases that can happen:

- **First Case**: the input $x^{(p)}$ matches any of the X rules in the rule base, and in this situation the system follows steps explained in **Case1**.
- **Second Case**: the input $x^{(p)}$ does not match any of the X rules in the rule base, then in this situation the system follows steps explained in **Case 2**.

Steps of **Case1** and **Case2** will be described in the following section.

➔ **Case 1: Input matches one of the existing rules:**

This case the input pattern $x^{(p)}$ can generate one or more than one rule. If the input pattern $x^{(p)}$ generates one rule and this rule matches any rule in the rule base then only the predicted class for this input pattern $x^{(p)}$ is the consequence class. However, if the input pattern $x^{(p)}$ generates more than one rule, then in this situation there is a possibility for conflict in the consequences of these rules.

Any of the generated rules by the input pattern $x^{(p)}$ can match different rules in the rule base, and any of these rules in the rule base may have a different consequence class. In this case we will need to choose one of these classes to be the predict class for the input pattern $x^{(p)}$. To do this we need to calculate a vote for each class as follows [2]:

$$\bar{z}Class_h\left(x^{(p)}\right) = \frac{\Sigma_{x\,\epsilon\, h}\overline{f^{J}}(x^{(p)}) * \overline{wd}(\tilde{A}_q \rightarrow C_q)}{max_{j\,\epsilon\, h}\overline{f^{J}}((x^{(p)}) * \overline{wd}(\tilde{A}_q \rightarrow C_q))} \tag{17}$$

$$\underline{z}Class_h\left(x^{(p)}\right) = \frac{\Sigma_{x\,\epsilon\, h}\underline{f^{j}}(x^{(p)}) * \underline{wd}(\tilde{A}_q \rightarrow C_q)}{max_{j\,\epsilon\, h}\underline{f^{j}}(x^{(p)}) * \underline{wd}(\tilde{A}_q \rightarrow C_q))} \tag{18}$$

After we calculate the upper $\bar{z}Class_h\left(x^{(p)}\right)$ and $\underline{z}Class_h\left(x^{(p)}\right)$ now we calculate the total vote strength for each of competitors classes as follows[2]:

$$zClass_h = \frac{\bar{z}Class_h\,(x^{(p)}) + \underline{z}Class_h\,(x^{(p)})}{2} \quad (19)$$

Now the class with the highest $zClass_h$ will be the winner class and it will be taken as predict class for the input pattern $x^{(p)}$.

➔ **Case 2: Input does not match any of the existing rules:**

This case the input pattern $x^{(p)}$ can generate more than one rule which could not match any existing rule in the rule base. In this situation we need to decide which one of the two classes could be the predicted class for the input pattern $x^{(p)}$. In order to resolve this conflict we let $MR\ (x^{(p)})$ be the set of rules generated by the input pattern $x^{(p)}$.

For any rule in $MR\ (x^{(p)})$ we find the closest rule in the rule base and then use the same steps as in **Case 1**. To find the closest rule for any rule in $MR\ (x^{(p)})$ in the rule base, we need to find the similarity (or distance) between any rule generated by the input pattern $x^{(p)}$ and each rule stored in the rule base. Then the rule with the highest similarity is selected to be the most similar rule. In order to calculate the similarity between one rule generated by the input pattern $x^{(p)}$and other rule stored in rule base we use the following equation[2]:

$$Similarity_{\ input\ r \leftrightarrow j} = ((1 - \left|\frac{vinput1r - vj1}{v1}\right|) * (1 - \left|\frac{vinput2r - vj2}{v2}\right|) * \ldots .* (1 - \left|\frac{vinputnr - vjn}{vn}\right|) \quad (20)$$

where

$vinput\ r = \ vinput\ 1r, vinput\ 2r, \ldots vinput\ nr)$ represent the linguistic labels that correspond to the rule generated by the input pattern $x^{(p)}$.

$v_j = (v_{j1,}\ v_{j2}\ \ldots\ v_{jn})$ represent the linguistic labels that corresponds to the rule stored in the rule base. Each of these linguistic labels could be decoded into an integer.

$v_1 \ldots v_n$ represent the number of linguistic labels representing each variable.

These steps lead to prediction of each rule generated by the input pattern $x^{(p)}$ where we will have the most similar rule in the rule base associated with similarity factor. Finally, to identify the final predicted class, we use the same steps as **Case 1** with multiplying each rule's "***weighted scaled dominance***" using (15) and (16) by its corresponding similarity factor.

V. *Experiments & Results*

The proposed model is evaluated using real-time financial data extracted from Sudanese banking sector. The data collection and analysis technique will be described in this section.

Data collection and Analysis:

The data set which was used to test the proposed model is data collected from Al-Shamal Islamic bank, Sudan. The data set is characterized as follows:

- Contains records dating back to the period between 2007 and 2017.
- Contains 101,257 records
- Contains 1,120 records catogized as defaults
- Contains 100,137 catogized as non defaults
- Collected from 23 bank branches distributed across Sudan.

To build the proposed model, the data was divided randomly to 70% used in learning phase, and 30% used for testing phase. The data set schema contains 12 parameters listed in Table 1 which shows parameters selected as inputs to the system with their description.

TABLE 1: THE SYSTEM INPUTS AND THEIR DESCRIPTION

Parameter Name	Description
AGE	costumer's age
SEX	costumer's gender
M_STATUS	costumer's marital status
DEP_CHILDREN	number of costumer's dependent children
Income	costumer's income per month
DEP_SPOUSES	number of costumer's dependent spouses
OCCUPATION	costumer's occupation
MONTH_EXP	costumer's average monthly expenditure
LIVE_COUN	costumer's live country
LIVE_CITY	costumer's live city
TOT_AMOUNT	total costumer's loan amount
CLASS	costumer's class type(default/not default)

We started by constructing the Type-2 fuzzy sets using equal spaced fuzzy sets. Then the Fuzzy C-mean clustering algorithm (FCM) was exercised to generate the Type-1 fuzzy sets, as these will be used in the following iteration of the experiment. The FCM output shown in Figure 7 represents distribution of the parameter age's data into three different clusters. Each of the clusters plays as a single fuzzy set i.e Young, Adult, and Old clusters. These sets can be aproximated in order to construct the corresponding Type-1 fuzzy set membership function as shown Figure 8.

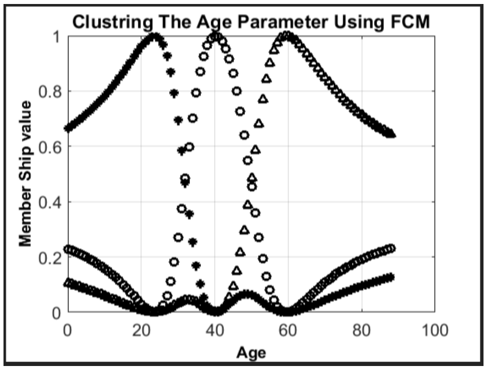

Figure 7: Output from FCM

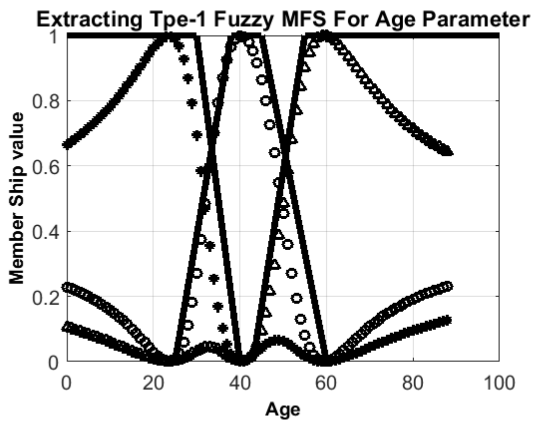

Figure 8: Sample of constructed Type-1 Fuzzy set

The generated Type-1 fuzzy sets were then tuned by extending the FOU 10%, 20%, and 30% consequently. This to ensure that we can setup three different groups of Type-2 fuzzy sets that were used throughout the experiments. The sample of constructed Type-2 fuzzy set with 10 % FOU is shown in Figure 9.

An average recall (AVG-Recall) method was used to measure the proposed model accuracy. The AVG-Recall could be calculated in a confusion matrix which displays information about predicted and actual classification done by a classifier[27]. This information is used to measure the classifier's performance.

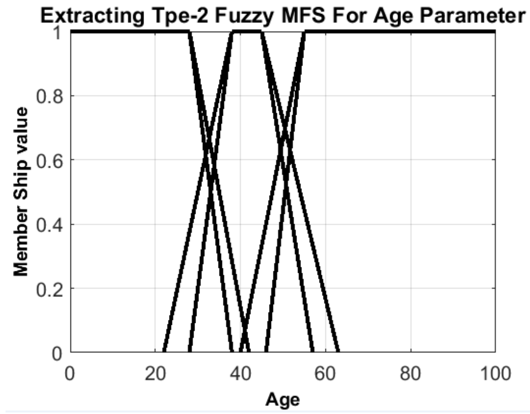

Figure 9: Sample of constructed type2 fuzzy set with 10% FOU

TABLE 2: CONFUSION MATRIX FOR BINARY CLASSIFICATION PROBLEM

	Actual Positive	Actual Negative
Positive Prediction	*True Positive(TP).*	*False Positive(FP)*
Negative Prediction	*False Negative(FN)*	*True Negative(TN)*
	Total Positive	Total Negative

If there exist an input item and two classes (positive and negative), then there would be four possible cases that can occur. These would be:

- The input item is positive and the classifier classifies it truly as positive and this case is known as **True Positive (TP)**.
- The input item is negative and the classifier classifies it as positive and this case is known as **False Positive (FP)**.
- The input item is positive and the classifier classifies it as negative and this case is known as **False Negative (FN)**.
- The input item is Negative and the classifier classifies it truly as negative and this case is known as **True Negative (TN)**.

Table 2 shows a confusion matrix for binary classification problem. From the information provided by the confusion matrix we can calculate ***Recall*** which is called sensitivity for both classes (positive and negative) as follow [26]:

$$\text{Recall Positive rate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{21}$$

$$\text{Recall Negative rate} = \frac{TN}{TN + FP} \tag{22}$$

Finally, now we can calculate the AVG-Recall which is used to measure the performance of the proposed model as follow[2]:

$$Avg\ Recall = \frac{Recall_{positive} + Recall_{Negative}}{2} \qquad (23)$$

In order to evaluate the proposed model four different experiments were conducted as follow:

1) Type-2 FLC using Type-2 fuzzy sets with equal space FOU.
2) Type-1 FLC using Type-1 fuzzy sets generated by FCM.
3) Type-2 FLC using Type-2 fuzzy sets generated by FCM with 10% FOU.
4) Type-2 FLC using Type-2 fuzzy sets generated by FCM with 20% FOU.

Tables 3 to 6 provide brief description for all conducted experiments. Table 3 shows confusion matrix for propsed Type-2 with 10% FOU. Table 4 display confusion matrix for propsed Type-1 system. Table 5 summarizes the results extracted using testing data. Table 6 shows the result of experiment that were extracted using the training data.

TABLE 3: CONFUSION MATRIX FOR PROPSED TYPE-2 WITH 10% FOU SYSTEM

	Actual Positive	Actual Negative
Positive Prediction	*TP =255.*	*FP=0*
Negative Prediction	*FN*= 135	*TN*= 30041
	Total Positive	Total Negative

TABLE 4: CONFUSION MATRIX FOR PROPSED TYPE-1 SYSTEM

	Actual Positive	Actual Negative
Positive Prediction	***TP =273.***	***FP=9500***
Negative Prediction	***FN***= 117	***TN***= 20541
	Total Positive	Total Negative

From Table 5, it can be noticed that the Type-2 fuzzy based system using FCM with 10% FOU outperform Type-1 fuzzy based system using FCM. The improvement is computed as 20.66%.

TABLE 5: TESTING DATA RESULTS SUMMARY

Exp #	Model type	FCM	FOU	Avg-Recall
1	Type-2	No	equal	0.798
2	Type-1	Yes	0	0.692
3	Type-2	Yes	10%	0.835
4	Type-2	Yes	20%	0.828

TABLE 6: TRAINING DATA RESULTS SUMMARY

Exp #	Model type	FCM	FOU	Avg-Recall
1	Type-2	No	equal	0.994
2	Type-1	Yes	0	0.913
3	Type-2	Yes	10%	0.995
4	Type-2	Yes	20%	0.967

The rule base that generated contain 8214 rules. Table (7) shows example of extracted rules by the proposed model. This provides an insight on the model operation as the main advantage provided by white box models. By analyzing these rules, the decision maker can reduce the potential risks that can face the organization as well as protecting customers form defaulting through advising in accordance to the analyzed information.

TABLE 7: EXAMPLE OF EXTRACTED RULE BY PROPOSED MODEL

N	*Rule*
R1	*if age is **Young** & sex is **Male** & marital_Status is **Married** & no_Dep_Child is **Mid** & income is **Low** & no_Dep_Spouses is **Low** & accupation is **Basic** & Avg_Month_Exp is **High** & live_Country is **SD** & live_City is **Khartoum** & tot_Amoount is **Low** Then class is **default***
R2	*if age is **Young** & sex is **Male** & marital_Status is **Married** & no_Dep_Child is **High** & income is **Low** & no_Dep_Spouses is **Low** & accupation is **HigherEducation** & Avg_Month_Exp is **Low** & live_Country is **SD** & live_City is **Khartoum** & tot_Amoount is **High** Then class is **default***

If **R2** customer is evaluated as shown in Table (7), one can notice that customer is catogrized with "HIGH" number of children, "LOW" income, and try to borrow a "HIGH" amount. This results in a logical reasoning why this customer had defaulted.

For example if we analyze customer **R1** in Table (7) we can simply find that this customer is "Young" and customer accupation is catogrized as “BASIC”. On one hand, this indicates that rational experience percentage loss, and on the other hand customer income is “LOW” and month_exp is “HIGH”. Therefore, all of these indicators conclude that this customer is a potential for a default.

VI. Conclusions

A Type-2 Fuzzy logic model is proposed for decision support. The model is validated with real financial data extracted from Sudanese banking sector. The model has been able to identify financial default in the data and provided factors led to the decisions. The proposed system resulted in transparent outputs which could be easily understood, analyzed and augmented by the human stakeholders. The model has shown excellent average recall of 83.5%, which outperformed its Type-1 counterpart by 20.66%. Furthermore, the rule base which had been extracted by the proposed model provided a good tool to help decision makers analyse customer data and understand reasons behind model predictions. This an attractive feature to the organization as well as to the customer avoiding default situations. Such advantage cannot be provided by using black box models.

REFERENCES:

[1] P. L.Swagel, Cost of the Financial Crisis, http://www.pewtrusts.org/en/research-and-analysis/reports/2010/04/28/the-impact-of-the-september-2008-economic-collapse, retrieved (28 April 2010).

[2] D.Bernardo ,H. Hagras, and E. Tsang, (2013) “A genetic type-2 fuzzy logic based system for the generation of summarized linguistic predictive models for financial applications”, Soft Computing, **17,** 2185-2201.

[3] A. Salih and H. Hagras, (2018) "Towards a Type-2 Fuzzy Logic Based System for Decision Support to Minimize Financial Default in Banking Sector", In: Proceeding of 10th Computer Science and Electronic Engineering (CEEC), Colchester, United Kingdom, pp:46-49.

[4] Daniel Martin, (1977)“Early warning of bank failure: A logit regression approach”, Journal of banking & finance, **1**, pp: 249-276.

[5] J. A. Ohlson, (1980). “Financial ratios and the probabilistic prediction of bankruptcy”, Journal of accounting research 18, 109-131.

[6] W. J. Banks, and P. L. Abad (1994). “On the performance of linear programming heuristics applied on a quadratic transformation in the classification problem*”.European Journal of Operational Research*“ 72, 23-28.

[7] C. Robert West,(1985) ”A factor-analytic approach to bank condition”, Journal of Banking & Finance 9,253-266.

[8] H. Jae Min and Young-Chan Lee(2005)” Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters”, Expert systems with applications 28, 603-614.

[9] K. Y. Tam,(1991) “Neural network models and the prediction of bank bankruptcy”, Omega 19 ,429-445.

[10] M. Linda Salchenberger, Mine Cinar, and Nicholas Lash,(1992) “Neural networks: A new tool for predicting thrift failures”, Decision Sciences23, 899-916.

[11] S. Piramuthu, H. Ragavan, and M. Shaw, (1998) ”Using feature construction to improve the performance of neural networks”, Management Science 44, 416-430.

[12] F. Barboza, H. Kimura, E. Altman (2017),” Machine learning models and bankruptcy prediction”, Expert Systems with Application 83, 405-417

[13] Karen Ketler, (1993) “Case-based reasoning: An introduction”, Expert systems with Applications 6,3-8.

[14] Cheol-Soo Park, and Ingoo Han, (2002)” A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction”, Expert Systems with Applications 23, 255-264.

[15] P. Ravi Kumar and V. Ravi, (2007) “Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review”, European journal of operational research. 180**,** 1-28.

[16] Halina Frydman, Edward I. Altman, and Duen-Li KAO, (1985) “ Introducing recursive partitioning for financial classification: the case of financial distress”, The Journal of Finance 40,:269-291.

[17] L. A. Zadeh, (1978) “Fuzzy sets as a basis for a theory of possibility”, Fuzzy Sets and Systems 1:3-28.

[18] [18] J. De Andrés, M Landajo, and P. Lorca (2005) “Forecasting business profitability by using classification techniques: A comparative analysis based on a Spanish case”, European Journal of Operational Research 167,518-542.

[19] S. Michael, D. Georgios, M. Nikolaos, and Z. Constantin, (1999). “A Fuzzy Knowledge-Based Decision Aiding Method for the Assessment of Financial Risks: The Case of Corporate Bankruptcy Prediction”, In: Proc. European

Symposium on Intelligent Techniques, Crete, Greece.

[20] N. N. Karnik and J. M. Mendel, (2001). "Operations on type-2 fuzzy sets", Fuzzy sets and systems 122,327-348.

[21] N.N. Karnik, and J.M. Mendel, (1998), "Introduction to type-2 fuzzy logic systems,in :Proc.1998 IEEE *International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence,* pp. 915-920, Anchorage, AK, USA..

[22] A. Bilgin, H. Hagras, J. Helvert, and D. Alghazzawi (2016). "A Linear General Type-2 Fuzzy Logic Based Computing With Words Approach for Realising an Ambient Intelligent Platform for Cooking Recipes Recommendation", IEEE Transactions on Fuzzy Systems, 24, 306-326.

[23] A. Starkey, H. Hagras, S. Shakya, and G. Owusu (2016). "A Multi-Objective Genetic Type-2 Fuzzy Logic Based System for Mobile Field Workforce Area Optimization", Journal of Information Sceinces 333,390-411.

[24] C. James Bezdek, Robert Ehrlich, William Full(1984), "FCM: The fuzzy c-means clustering algorithm, Computers & Geosciences10,191-203.

[25] S. Ghosh and S. Kumar(2013,) "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms," International Journal of Advanced Computer Science and Applications 4,35-39.

[26] H. Ishibuchi, and T. Yamamoto (2005). "Rule Weight Specification in Fuzzy Rule-Based Classification Systems", IEEE Transactions On Fuzzy Systems 13, 428-435.

[27] R. Kohavi, and F.P rovost (1998), "Glossary of terms", Edited for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process 30.