

Hybrid Reliability-Similarity-Based Approach for Supervised Machine Learning

Walid Cherif

Abstract—Data mining has, over recent years, seen big advances because of the spread of internet, which generates everyday a tremendous volume of data, and also the immense advances in technologies which facilitate the analysis of these data. In particular, classification techniques are a subdomain of Data Mining which determines in which group each data instance is related within a given dataset. It is used to classify data into different classes according to desired criteria. Generally, a classification technique is either statistical or machine learning. Each type of these techniques has its own limits. Nowadays, current data are becoming increasingly heterogeneous; consequently, current classification techniques are encountering many difficulties. This paper defines new measure functions to quantify the resemblance between instances and then combines them in a new approach which is different from actual algorithms by its reliability computations. Results of the proposed approach exceeded most common classification techniques with an f-measure exceeding 97% on the IRIS Dataset.

Keywords—Data mining, knowledge discovery, machine learning, similarity measurement, supervised classification.

I. INTRODUCTION

DATA Mining is a recent field at the cross junction of statistics and artificial intelligence, its purpose is the discovery of structures and information in large data sets. It consists of analyzing this profusion by using different algorithms which will explore the most infinite relation between these data and then deliver nontrivial, previously unknown and potentially useful information. Data mining was presented in 2001 by the Massachusetts Institute of Technology as one of the 10 emerging technologies that are changing the world in the 21st century [1].

In this dynamic environment, characterized by both technological advances on the one hand and the huge quantity of generated data on the second, analytical needs have also evolved, requiring the development of advanced algorithms and tools increasingly powerful which should be capable of adapting the contextual information to the types of data processed [2].

In general, Data Mining consists of solving a problem related to the considered data. There are four main issues addressed by this step:

- Association: Formally, it consists in discovering the most frequent relationships between attributes [3].
- Prediction: it seeks to estimate the value of a target attribute based on the values of the available attributes.

Walid Cherif is with the Laboratory SI2M, National Institute of Statistics and Applied Economics, B.P. 6217, Rabat, Morocco (e-mail: chrif.walid@gmail.com).

Depending on the type of the target variable, we distinguish: classification when predicting discrete values and regression when predicting numerical values [4].

- Grouping: it aims to generating a partition, consisting of a set of clusters where the data of the same cluster are very similar [5].
- Detection of anomalies: it detects aberrant behaviors in supposedly homogeneous data [6].

These problems generate two types of Data Mining model: predictive (supervised) and exploratory (unsupervised). Supervised learning consists of two major steps: training and testing [7].

The problem of classification is central in the majority of these applications. By its simplicity, it represents, in a way, the cornerstone of many treatments and gives rise to a variety of applications. It concerns business [8], medicine [9], finance [10], text analysis [11], and image analysis [12], etc. As a result, it has generated the most data mining work, so a large number of algorithms in the literature are devoted to the classification of textual data, which makes their complete presentation impossible. The most widely used supervised algorithms in the literature are Naïve Bayes (NB) [13], Support Vector Machines (SVM) [14], k-Nearest Neighbors (KNN) [15] and Artificial Neural Networks (ANN) [16].

The goal of this study is to build up a new accurate statistical approach (CSBS) that supports different types of data, and that learns to classify from most reliable attributes more accurately.

The rest of this paper is structured as follows: Section II summarizes main works on supervised machine learning. Section III meticulously details the modeling process, as it is crucial in the build of the CSBS decision function; and in Section IV, the results of four most known classifiers are compared to the CSBS. Finally, the last section concludes this work.

II. BACKGROUND

Until the end of the 1990s, the dominant approach to construct prediction models in general, and classification in particular, was based on a knowledge engineering approach. An expert system was constructed with a set of rules defined manually in order to emulate the decision-making ability of a human expert [17]. Expert systems were among the first truly successful forms of artificial intelligence [18]. However, the development of artificial intelligence techniques and computer performances has allowed the process of classification from another perspective. Rather than resorting to experts to establish the rules, the alternative came from a subdomain of

artificial intelligence that developed strongly during the 1990s: machine learning that allows the system to learn by itself to accomplish this task [19]. Since then, several algorithms have succeeded:

A. K-Nearest Neighbors

The algorithm KNN is among the simplest artificial learning algorithms based on similarities. The basic idea when classifying a given instance is to vote its nearest neighbors in the sense of a predefined distance. The class of the new instance is then determined by the majority among the KNN.

The performance of KNN depends largely on two factors: the value of k and the measure used [20]. Generally, it becomes poor when the features are too many, or when data are very heterogeneous.

B. Naïve Bayes

NB classifier is based on Bayes' theorem [21]. The basic idea is to estimate the probabilities of membership of the samples to their classes. The naive part of this model is the assumption that all variables are independent. The simplicity of this assumption makes the computation of NB classifier far more efficient [22].

C. Artificial Neural Network

This method is very schematically inspired from the function of biological neurons; hence its name. The neurons receive the signals (electrical impulses) through highly branched extensions of their cellular body (dendrites) and they send the information through long extensions (axons) [23].

The algorithm ANN learns a model by means of a feed-forward neural network trained by a back propagation algorithm. A neuron is primarily a mathematical operator. It performs a weighted sum, followed by a nonlinear function. This function must be bounded, continuous and differentiable; the most frequently used ones are sigmoid functions [24].

D. Support Vector Machines

The foundations of SVM originated from early concepts developed by Cortes and Vapnik [14], this method has proven to be very robust for general classification and regression [25].

SVM are based on two key ideas: the notion of maximum margin and the concept of kernel function. In linear classification, they create a hyper plan that separates the data into two sets with a maximum margin [26]. For the cases where the data are not linearly separable, they map the data representation space into an area of larger dimension in which it is probable that there is a linear separator.

Further works combined two or more of the previous algorithms. Machhale et al. proposed a classification system to recognize normal and abnormal MRI brain images by combining SVM and KNN [27]. Another work proposed a hybrid method based on ANN and KNN in a scheme intended for cloud classification [28]. In a work on opinion mining, noting that some unclassified instances by SVM could be done by KNN, authors have hybridized the SVM by KNN, thus raising the overall performance [29].

The optimization of previous algorithms also aroused the

interest of researchers, Zhang et al. proposed a hybrid method applying SVM with an optimization of its parameters in an application on fault error diagnosis [30]. ANN and KNN algorithms have also been optimized to meet the need for an accurate classification [31], [32]. This diversity of algorithms is now applied in different domains:

The most remarkable interest was devoted to Customer relationship management (CRM) [33]. It comprised a set of processes and enabled systems supporting a business strategy to build long term, profitable relationships with specific customers [34]. The choices of data mining techniques should be based on the data characteristics and business requirements [35].

The field of healthcare also benefited from data mining algorithms; several works in the literature have applied supervised and unsupervised classification techniques whether to diagnose tumors [36], or predict the causes of a given disease [37], and many other clinical applications [38].

Classification techniques have also been applied to textual data, whether for automatic categorization of texts [39], or sentiment analysis [40], or simply for information retrieval [41].

Nowadays, these algorithms are applied in many other fields, and the need to construct more reliable approaches is becoming more and more indispensable [42].

III. THE PROPOSED APPROACH

In this section, the main problems faced by previous machine learning algorithms are highlighted, and new measures are defined in order to build up the new accurate approach we called CSBS.

The first faced problem while computing similarities for many attributes is the normalization:

Let's consider the two examples: E_1 which belongs to the first class a , and E_2 which belongs to the second class b :

TABLE I
EXAMPLE OF NON-STANDARDIZED DATA

	x_1	x_2	x_3	x_4	Y
E_1	3000	30	36	12	a
E_2	1200	55	64	21	b
	x_1	x_2	x_3	x_4	
E_3	2800	55	62	20	

TABLE II
DISTANCES TO CLASSES

Distance \ class	a	b
Manhattan	64.75	400.75
Euclidean	101.69	800.00

When the distance is calculated directly, the impact of the order of magnitude is so important. For example, the new example E_3 has very close values to the example E_2 on the three attributes: x_2 , x_3 and x_4 ; but the order of magnitude of x_1 (10^3) makes the entire distance very far, favoring thus the example of the class a which has relatively close value only for x_1 : In order to overcome this problem, a normalization is adopted for each value e_{ij} of the attribute x_j for the example E_i

as:

$$e_{ij} \leftarrow \frac{e_{ij} - \min_j(e_{ij})}{\max_j(e_{ij}) - \min_j(e_{ij})} \quad (1)$$

The second problem concerns similarities. To highlight it, let's consider the normalized training dataset:

TABLE III
Data Standardization

	x_1	x_2	x_3	x_4	Y
E_1	0.2	0	0.8	1	a
E_2	0.2	0.25	0.82	0.9	a
E_3	0.1	0.6	0.85	0.9	a
E_4	1	0.5	1	0.45	b
E_5	0.1	0.3	0.2	0.7	b
E_6	0.5	1	0.25	0.4	b
E_7	0.2	0.2	0.1	0	c
E_8	0	0.2	0	0.25	c
E_9	0.8	0.2	0.83	0	c

The goal is to build up a model that would be able to classify correctly any given new observation:

TABLE IV
TEST OBSERVATION

	x_1	x_2	x_3	x_4
E_{10}	0.3	0	0.83	0.9

For each class from {a,b,c}, the following characteristics are retained:

- The number of examples: n_k
- And for each attribute:
- The amplitude: A_{jk}
- The own amplitude: A^*_{jk}
- The minimal value: $\min_{jk}(e_{ij})$
- The maximal value: $\max_{jk}(e_{ij})$
- The center of the class: c_{jk}

For example, for the class a, the attribute x_1 has the following characteristics:

- The number of examples: $n_a = 3$
- The minimal value: 0.1
- The maximal value: 0.2
- The center of the class: $c_{1a} = \frac{0.1+0.2+0.2}{3} = 0.17$
- The amplitude: $A_{1a} = 0.2 - 0.1 = 0.1$
- The own amplitude: $A^*_{1a} = 0$ because the class c also has values in this interval (Fig. 1).

In what follows, different sub measures of similarities are discussed:

A. Equal Values n_k

While classifying E_{10} , let us begin with the attribute x_4 : it has the same value as two examples from the class a (66%). Similarly, x_2 has one equal value belonging to the same class a.

By adding the rest of attributes, the class of E_{10} is the one having most equal values on all attributes. This measure is very valuable especially for discrete attributes, and it is used

by Decision Tree algorithms to compute the information gain. However, being limited to this measure alone is very insufficient as it can be clearly seen for the attribute x_3 for instance, which has one equal value from the class c; but the two other examples are so far, and contrariwise, the values of the class a are very close to it; hence, the need to consider the concentration of values around the center of the class (especially for continuous attributes).

B. Distance to Center d_c

To overcome the aforementioned problem of concentration of values, distances to centers can be envisaged.

The attribute x_3 which was close to the values of the class a without being exactly equal to, has now the nearest value to its center:

TABLE V
DISTANCES TO CENTERS

	center of a	center of b	center of c
Distance to x_3	0.01	0.35	0.52

C. Amplitude a_{jk}

The computation of the amplitude is used to express the reliability of each attribute:

Let's reconsider the example of Table III, for a given example E_{11} , the attribute x_1 defines three possible cases according to its value:

- If $e_{11,1} \in [0.8 ; 1]$: E_{11} has a big chance to belong to the class b as it is the only class which takes values between 0.8 and 1.
- If $e_{11,1} \in [0.2 ; 0.8]$: It is impossible to know exactly if E_{11} belongs to the class b or c, as both of them include examples with values in this interval.
- If $e_{11,1} \in [0.1 ; 0.2]$: All classes have values in this interval; hence, the need to compute a new measure of similarity.
- If $e_{11,1} \in [0 ; 0.1]$: E_{11} has a big chance to belong to the class c.

D. Own Amplitude a^*_{jk} :

The own amplitude of a given attribute is introduced to predict if this latter is reliable relatively to other attributes (see Fig. 1). It eliminates the intervals with uncertain values.

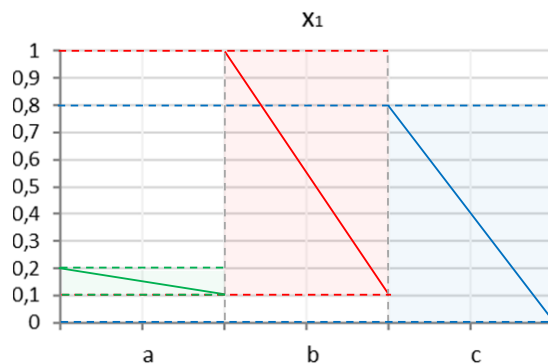


Fig. 1 The own amplitude of the attribute X1

E. Decision Function

To classify a given instance E_i , for each class k , the proposed approach computes the likelihood, which is defined as:

$$\xi_{lk} = \frac{1}{c_k} \times \sum_{j=1}^p p_{jk} \times \frac{N_j^{k^*} + a_j^{k^*} + a_j^{k^*} - d(x_{lj}, \bar{x}_j^k)}{N_j + 2a_j^{k^*} + \varepsilon_0} \quad (2)$$

where: a_j^k denotes the amplitude of the interval of values of k , $a_j^{k^*}$ denotes the own amplitude of the interval of values of k , N_j^k is the number of instances of the class k having a value equal to X_i on the attribute x_j , N_j is the total number of instances having a value equal to E_i on x_j , $d(x_{lj}, \bar{x}_j^k)$ denotes the considered distance between E_i and the center \bar{x}_j^k , p_{jk} is the coefficient of reliability on x_j to predict the class k , c_k is the number of instances of the class k , and, ε_0 is a very small positive value added to avoid a zero denominator (where all instances take the same value for x_j , and this latter is different from that of E_i .) The numerator remains zero for this case. Finally, the class of E_i is:

$$Y_i = \operatorname{argmax}_k \xi_{lk} \quad (3)$$

A threshold δ is chosen to eliminate non-significant attributes (attributes with smallest reliability coefficients). Such a threshold greatly increases the precision of the classification since it relies solely on the attributes which separates the classes, but increasing it leads to over-classification.

IV. RESULTS AND DISCUSSION

To evaluate the performance of the proposed approach, the effectiveness is compared to most common approaches of classification: k-NN, ANN, SVM, NB, on IRIS dataset which is a classic example of datasets for classification [42].

A. Computation

The dataset was divided into five subsets of 30 instances each (10 instances per class: I. setosa; I. versicolor; I. virginica).

At each iteration, four subsets are considered for training and the fifth for the test, the process is thus repeated for the five subsets, and the average f-measure is retained.

The F-Measure used for evaluation is written as:

$$\text{Recall} = \frac{a}{a+c}; \text{Precision} = \frac{a}{a+b}; \text{FM} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

where a : is the number of instances correctly classified, b : the number of false positives, and c : is the number of false negatives.

In what follows, the results of classification of the proposed algorithm CSBS and four other known algorithms:

B. K-Nearest Neighbors

Fig. 2 shows the f-measure of KNN algorithm for different

values of k . The optimal performance concerns $k = 3$, and the f-measures of the five subsets are: 100%, 96.82%, 93.89%, 96.82% and 96.82%, with a total of five instances erroneously classified.

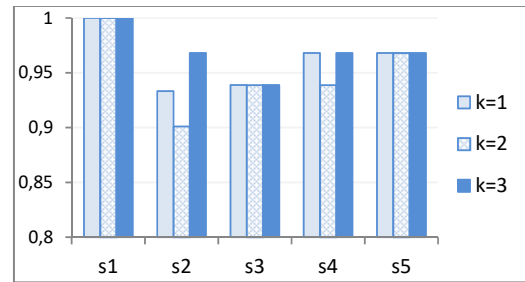


Fig. 2 f-measure of KNN algorithm for different values of k

C. Naïve Bayes

In Fig. 3, the NB algorithm generates the following results on the five subsets: 100%, 93.89%, 93.89%, 96.82% and 93.33%. The average f-measure is: 95.58%.

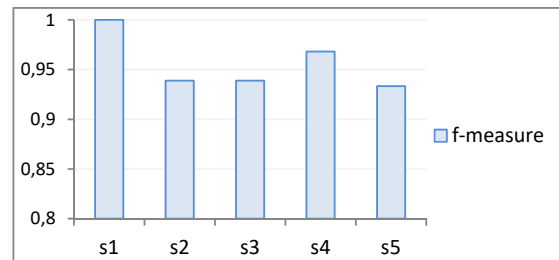


Fig. 3 f-measure of NB algorithm

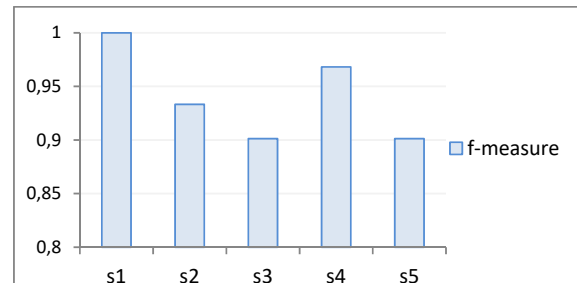


Fig. 4 f-measure of SVM algorithm with a RBF kernel

D. Artificial Neural Network

Fig. 4 summarizes the f-measure of ANN on the five subsets: 100%, 93.33%, 91.14%, 100% and 100%.

E. Support Vector Machines

In Fig. 5, the results concern SVM with RBF kernel [43] $\gamma = 0$, $C = 0$ and $\varepsilon = 10^{-3}$. On the five subsets, the reported f-measures are: 100%, 93.33%, 90.12%, 96.82% and 90.12%.

F. CSBS

The f-measure of CSBS algorithm showed a stability on the five subsets: 100%, 96.82%, 93.89%, 100%, 96.82%. None of the subsets generated more than three erroneous instances. The average f-measure is the highest among previous

algorithms: 97.50%.

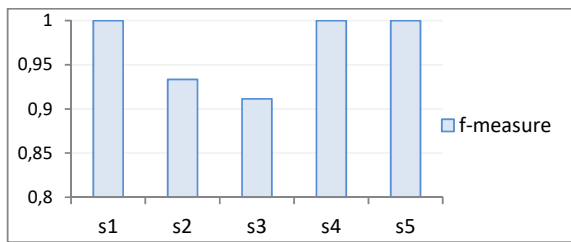


Fig. 5 f-measure of ANN algorithm

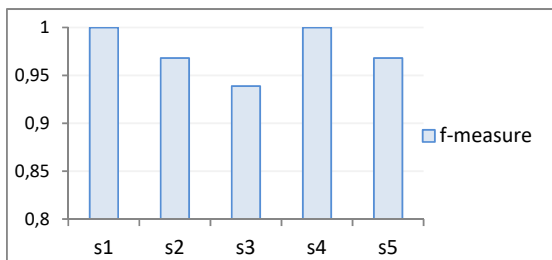


Fig. 6 f-measure of CSBS algorithm

The overall results are shown in Fig. 7 in terms of f-measure: the algorithm CSBS slightly exceeds KNN ($k = 3$) with the highest average f-measure: 97.5%. ANN returned the third best f-measure: 96.89%, and finally, NB and SVM (with a RBF kernel function) returned the weakest f-measures: respectively 95.58% and 94.07%. Moreover, the f-measure of CSBS algorithm is relatively stable on the five subsets. The weakest reported f-measure was 93.89%, and it concerns the training set with the lowest reliability coefficients, which means that when the attributes are sufficiently reliable, the CSBS algorithm outperforms other algorithms; But when this is not the case, some algorithms may offer better results, such as ANN on the subset s5. Future improvements can then be considered to hybridize CSBS with other supervised classification algorithms, for example with ANN, in order to make it even more efficient.

The outperforming of the CSBS algorithm is even more remarkable when the data type is more heterogeneous (for example on a sentiment analysis dataset).

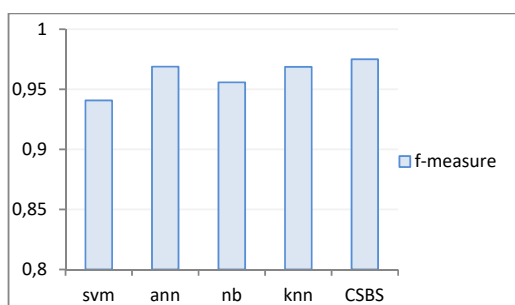


Fig. 7 Average f-measure of the considered classifiers

V. CONCLUSION AND PERSPECTIVES

Data mining is a fast expanding field with many new research results and new approaches developed. Applications of data mining techniques affect almost every aspect of our daily lives: industry, medicine, and finance, etc., and it has attracted the attention of practitioners and academics.

Nowadays, as the world is witnessing the increasingly complex needs of big data, classic algorithms are pushed to their limits, and new accurate machine learning approaches are necessary to deal with the variety of available data. In this sense, this paper introduces a new supervised machine learning approach that aims to overcome main drawbacks of classic classification models. This algorithm, called CSBS, selects most reliable attributes for each class, and then combines its reliability coefficients and special similarity measures to classify any given instance. A standard dataset is used to evaluate the proposed algorithm, and the results indicate that it outperforms most known classification algorithms, namely KNN, NB, ANN and SVM with a f-measure exceeding 97%. Afterwards, an analysis of the final results highlighted that CSBS can be further optimized by hybridization with ANN classifier. This optimization is the subject of our future researches.

REFERENCES

- [1] Stéphane, T. (2012). Data Mining et statistique décisionnelle: L'intelligence des données. Éditions Technip.
- [2] Drakos, N., Mann, J., Cain, M. W., Andrews, W., Knox, R. E., Valdes, R., ... & Harris, K. (2008). Hype Cycle for Social Software, 2008.
- [3] Pattanapruteep, O., McEvoy, M., Attia, J., & Thakkestian, A. (2017). Evaluation of rational nonsteroidal anti-inflammatory drugs and gastro-protective agents use; association rule data mining using outpatient prescription patterns. BMC Medical Informatics and Decision Making, 17(1), 96.
- [4] Marinakos, G., & Daskalaki, S. (2016). Viability prediction for retail business units using data mining techniques: a practical application in the Greek pharmaceutical sector. International Journal of Computational Economics and Econometrics, 6(1), 1-12.
- [5] Luam, P., Lin, H. W., Chiu, Y. P., Shyu, Y. L., & Lee, P. C. (2016). The Categorising Characteristics of Facebook Pages: Using the K-Means Grouping Method. International Journal of Business and Management, 11(2), 60.
- [6] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.
- [7] Lad, H., & Mehta, M. A. (2017). Feature Based Object Mining and Tagging Algorithm for Digital Images. In Proceedings of International Conference on Communication and Networks (pp. 345-352). Springer, Singapore.
- [8] Shmueli, G., & Lichtendahl Jr, K. C. (2017). Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. John Wiley & Sons.
- [9] Kumari, M., & Godara, S. (2011). Comparative study of data mining classification methods in cardiovascular disease prediction, 2(2), pp. 304-308.
- [10] Lin, C. C., Chen, C. S., & Chen, A. P. (2017). Using Intelligent Computing and Data Stream Mining for Behavioral Finance associated with Market Profile and Financial Physics. Applied Soft Computing.
- [11] Cherif, W., Madani, A., & Kissi, M. (2016, October). A combination of low-level light stemming and support vector machines for the classification of Arabic opinions. In Intelligent Systems: Theories and Applications (SITA), 2016 11th International Conference on (pp. 1-5). IEEE.
- [12] Slavenas, M., Rodriguez, P., Craig, A., Wuerffel, E., & Will, J. (2016, July). Image Analysis and Infrastructure Support for Data Mining the Farm Security Administration: Office of War Information Photography

- Collection. In Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale (p. 1). ACM.
- [13] Patil, T. R., & Shrekar, S. S. (2013). Performance analysis of Naive Bayes and J48 classification algorithm for data classification. *International Journal of Computer Science and Applications*, 6(2), 256-261.
- [14] Vapnik V (1995) *The nature of statistical learning theory*. Springer, New York.
- [15] Hwang, W. J., & Wen, K. W. (1998). Fast KNN classification algorithm based on partial distance search. *Electronics letters*, 34(21), 2062-2063.
- [16] Wasserman, P. D. (1993). *Advanced methods in neural computing*. John Wiley & Sons, Inc.
- [17] Jackson, P. *Introduction to expert systems*. 1998. Harlow, UK: Addison Wesley, 3.
- [18] Russell, S. J., & Norvig, P. (2002). *Artificial intelligence: a modern approach* (International Edition).
- [19] Weiss, S. M., & Kapouleas, I. (1990). An empirical comparison of pattern recognition, neural nets and machine learning classification methods. *Readings in machine learning*, 177-183.
- [20] Goldberger, J., Hinton, G. E., Roweis, S. T., & Salakhutdinov, R. R. (2005). Neighbourhood components analysis. In *Advances in neural information processing systems* (pp. 513-520).
- [21] Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (1993). *"Probability and statistics for engineers and scientists"* (Vol. 5). New York: Macmillan.
- [22] Lewis, D. D. (1998, April). Naive (Bayes) at forty: The independence assumption in information retrieval. In *European conference on machine learning* (pp. 4-15). Springer, Berlin, Heidelberg.
- [23] Dubois, D. (2011). *Intelligence Naturelle et Intelligence Artificielle*. *Acta Europaea Systemica*, 1, 1-10.
- [24] Wang, S. C. (2003). *Artificial neural network*. In *Interdisciplinary Computing in Java Programming* (pp. 81-100). Springer US.
- [25] Vapnik V. (2000). *The nature of statistical learning theory*. Springer Science & Business Media.
- [26] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Statistics and computing* 2004; 14-3. p.199-222.
- [27] Machhale, K., Nandpuru, H. B., Kapur, V., & Kosta, L. (2015, May). MRI brain cancer classification using hybrid classifier (SVM-KNN). In *Industrial Instrumentation and Control (IIC), 2015 International Conference on* (pp. 60-65). IEEE.
- [28] Xia, M., Lu, W., Yang, J., Ma, Y., Yao, W., & Zheng, Z. (2015). A hybrid method based on extreme learning machine and k-nearest neighbor for cloud classification of ground-based visible cloud image. *Neurocomputing*, 160, 238-249.
- [29] Cherif, W., Madani, A., & Kissi, M. (2016). A hybrid optimal weighting scheme and machine learning for rendering sentiments in tweets. *International Journal of Intelligent Engineering Informatics*, 4(3-4), 322-339.
- [30] Zhang, X., Qiu, D., & Chen, F. (2015). Support vector machine with parameter optimization by a novel hybrid method and its application to fault diagnosis. *Neurocomputing*, 149, 641-651.
- [31] Madić, M., Radovanović, M., Manić, M., & Trajanović, M. (2014). Optimization of ANN models using different optimization methods for improving CO2 laser cut quality characteristics. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 36(1), 91-99.
- [32] Mathieu-Dupas, E. (2010). Algorithme des k plus proches voisins pondérés et application en diagnostic. In *42èmes Journées de Statistique*.
- [33] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2), 2592-2602.
- [34] Ling, R., & Yen, D. C. (2001). Customer relationship management: An analysis framework and implementation strategies. *Journal of computer information systems*, 41(3), 82-97.
- [35] Giraud-Carrier, C., & Povel, O. (2003). Characterising data mining software. *Intelligent Data Analysis*, 7(3), 181-192.
- [36] Ma, J., Yin, S. Y., He, R. Q., Lu, Y. K., Xie, T. T., Shi, Q., ... & Gang, C. (2017). Significant target genes and signaling for miR-34a in gastrointestinal stromal tumors: a study of GEO-data mining and bioinformatics approaches. *international journal of clinical and experimental pathology*, 10(3), 2539-2553.
- [37] Abdar, M., Kalhori, S. R. N., Sutikno, T., Subroto, I. M. I., & Arji, G. (2015). Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. *International Journal of Electrical and Computer Engineering (IJECE)*, 5(6), 1569-1576.
- [38] Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: current issues and guidelines. *International journal of medical informatics*, 77(2), 81-97.
- [39] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *arXiv preprint arXiv:1707.02919*.
- [40] Cherif, W., Madani, A., & Kissi, M. (2015). Towards an efficient opinion measurement in Arabic comments. *Procedia Computer Science*, 73, 122-129.
- [41] Krallinger, M., Rabal, O., Lourenço, A., Oyarzabal, J., & Valencia, A. (2017). *Information Retrieval and Text Mining Technologies for Chemistry*. Chemical Reviews.
- [42] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of human genetics*, 7(2), 179-188.
- [43] Kuo, B. C., Ho, H. H., Li, C. H., Hung, C. C., & Taur, J. S. (2014). A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1), 317-326.