information

MDPI

Article

Text Classification Using Intuitionistic Fuzzy Set Measures—An Evaluation Study

George K. Sidiropoulos, **Nikolaos Diamianos**, **Kyriakos D. Apostolidis** and **George A. Papakostas** *

MLV Research Group, Department of Computer Science, International Hellenic University, 65404 Kavala, Greece; georsidi@cs.ihu.gr (G.K.S.); nidsmia@teiemt.gr (N.D.); kyriapos1@cs.ihu.gr (K.D.A.)
* Correspondence: gpapak@cs.ihu.gr

Abstract: A very important task of Natural Language Processing is text categorization (or text classification), which aims to automatically classify a document into categories. This kind of task includes numerous applications, such as sentiment analysis, language or intent detection, heavily used by social-/brand-monitoring tools, customer service, and the voice of customer, among others. Since the introduction of Fuzzy Set theory, its application has been tested in many fields, from bioinformatics to industrial and commercial use, as well as in cases with vague, incomplete, or imprecise data, highlighting its importance and usefulness in the fields. The most important aspect of the application of Fuzzy Set theory is the measures employed to calculate how similar or dissimilar two samples in a dataset are. In this study, we evaluate the performance of 43 similarity and 19 distance measures in the task of text document classification, using the widely used BBC News and BBC Sports benchmark datasets. Their performance is optimized through hyperparameter optimization techniques and evaluated via a leave-one-out cross-validation technique, presenting their performance using the accuracy, precision, recall, and F1-score metrics.

Keywords: Fuzzy Sets; Intuitionistic Fuzzy Sets; text classification; fsmpy

Citation: Sidiropoulos, G.K.; Damianos, N.; Apostolidis, K.D.; Papakostas, G.A. Text Classification Using Intuitionistic Fuzzy Set Measures—An Evaluation Study. *Information* **2022**, *13*, 235. https://doi.org/10.3390/info13050235

Academic Editors: Luis Martínez López, Lesheng Jin, Zhen-Song Chen and Humberto Bustince

Received: 3 April 2022
Accepted: 3 May 2022
Published: 5 May 2022

1. Introduction

Text classification is one of the most important tasks of Natural Language Processing (NLP) and Machine Learning (ML). The role of text classification (or text categorization) is to classify a document into predefined categories automatically. Texts contain much information that can be crucial for businesses; however, the manual extraction of such information can be very difficult and time consuming. For this reason, companies use text classifiers in order to manage the vast plethora of free texts in an effective way.

Text classification has experienced great interest in the literature, with methodologies focusing on various aspects of the classification and analysis process. One of the most important aims in the field is sentiment analysis [1,2], which determines if a text has a positive, negative, or neutral meaning. Other approaches aim to detect the language the text is written in [3,4] or even its intention, especially in customer conversations where they aim to understand the customer's purposes [5]. All these applications are mainly used in social media monitoring, brand monitoring, customer service, and the voice of the customer. Another common and important application of text classification is topic labeling, which aims to find the topic of a free text. A significant implementation of topic labeling is the organization or recommendation of articles according to their purposes [6].

The most common algorithms for text classification use machine learning and deep learning methods, which present very good results. Naive Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) are some of the most-used classifiers in machine learning for text classification. In deep learning, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are the most types of networks used to classify documents.

A very fitting mathematical theory for these kinds of problems is Fuzzy Set (FS) theory. FS theory, in contrast to classical set theory which assesses the belongingness of an element to a set in binary terms, allows for gradual assessments. Zadeh [7] proposed the theory of FSs in 1965 which introduced the degree of membership. FSs have become very useful, and their application has been considered in many real-world problems in order to deal with vague information. One of the most successful applications of Fuzzy Sets in the industrial and commercial field was achieved with fuzzy controllers [1]. Furthermore, they are widely used in problems with incomplete and imprecise information such as bioinformatics [8]. Another crucial application of FSs is the automobile industry using Mamdani's approach. Later, in 1986, Atanassov [9] introduced Intuitionistic Fuzzy Sets (IFSs), which are an improved version of FSs as they are able to solve more complex problems, introducing the non-membership function.

The application of FS theory in pattern recognition, medical diagnosis, image segmentation, or classification in general, has been considered and employed for many years now, and it is enabled by the use of Intuitionistic Fuzzy measures. Those measures aim to calculate the distance or similarity between the fuzzy patterns of a class and that of a new sample to determine to which the class it belongs. For this reason, the literature has focused extensively to develop such mathematical measures, which can be used to calculate the distance between patterns as best as possible, maximizing the distance between samples of different classes and minimizing it on similar samples. Moreover, FSs have the advantage of being able to handle uncertainties, which is a very common drawback of many approaches. Taking into account the important requirements of text classification combined with the usefulness of FS theory, the need to evaluate the existing IFS measures for their application in text document classification is required. For this purpose, in this work, we use a framework for text classification that was proposed by [10], with the aim of evaluating the application of IFS theory for text classification and the different IFS measures that have been proposed in the last 30 years, as well as to highlight the important aspects of the application of FS theory in the specific domain.

This paper contributes to the thorough study of text classification using Fuzzy Sets, with several different distance and similarity measures being tested on two famous text classification datasets. Moreover, the effectiveness of using FSs and IFSs for the specific problem is highlighted through a detailed comparison of the measures, using the framework mentioned above, as well as comparing their application with other, conventional, machine learning models.

The rest of the paper is organized as follows. Section 2 provides the related works on text classification using FS theory. Section 3 presents some information about Fuzzy and Intuitionistic Fuzzy Sets, along with the datasets, tools and evaluation protocol used to conduct the experiments. Section 4 shows the results obtained, and Section 5 discusses the results and compares them to those of conventional machine learning models. Finally, Section 6 concludes the paper.

2. Related Work

The theory of FS was proposed in 1965 by Zadeh [7] in order to deal with a large number of functional problems. In fact, FSs are the sweeping form of classical sets, which were carried out, basically, in control systems, as early as the 1980s. FSs allow to gradually assess if an element belongs to a set via a membership function, which values the belongingness in the real unit interval $[0, 1]$. This number reflects the degree of association of the element with a given set, with 0 meaning it does not belong to the set, whereas 1 means that it completely belongs to the set. This degree of association is called the membership value, and the non-association of a number in a set can be described by a value called the non-membership value, represented by 1 minus the membership value of the number.

A very common application of FSs is feature selection for text classification, where the incorporation of the fuzziness and fuzzy logic in feature selection has shown improvements.

A. Z. Bushra et al. [11] presented a feature selection methodology aiming to improve the existing issue of selecting overlapping features.

Fuzzy Sets and similarity measures have been exploited for text classification with good results [12]. Fuzzy Sets have been used in a Similarity-Based Concept Mining Model [13] for text categorization into pre-defined category groups. They have also been used for dimensionality reduction of the feature vector in the process of feature clustering in text classification [14]. Y. Jiang et. al [15] proposed a method that classified multi-labeled text using fuzzy similarity and K-Nearest Neighbors (KNNs), which showed very competitive results compared to other methods, avoiding the high computational cost of a KNN using a fuzzy similarity measure to group the patterns into clusters. Moreover, IFSs and similarity measures have been used for text categorization, yielding good results [10]. The proposed framework exploits statistical methods to represent a document as an IFS, "learning" then its IFS patterns. Szmidt and Kacprzyk [16] proposed another text classification algorithm, using IFSs, highlighting their importance for text classification due to their robustness in imbalanced classes.

More recently, the literature has shifted towards combining the theory of fuzzy logic and Fuzzy Sets with machine learning, especially Convolutional Neural Networks (CNNs), as they have shown great success in Natural Language Processing problems. For example, M. Bounabi et al. [17] exploited fuzzy logic and improved the performance of ML models such as Naive Bayes, Support Vector Machines, and Random Forest classifiers. On the other hand, B. Behera [18] combined a Fuzzy-Rough-Set-based robust nearest neighbors algorithm with a CNN-based feature extraction method, for text document classification and feature extraction, showing important improvements compared to other approaches and models. Moreover, S. Puri [19] presented two systems for text classification, called the Concept-Based Mining Model using a Threshold and the Fuzzy Similarity-based Concept Mining Model using Feature Clustering. These models perform the processes of cutting off, feature extraction, and feature selection during the preprocessing steps, showing high feature reduction and classification accuracy rates.

3. Materials and Methods

3.1. Important Definitions

Here, we describe the definitions of Fuzzy Sets and Intuitionistic Fuzzy Sets, as well as the definitions and constraints of a distance and similarity measure when applied in Fuzzy Set theory.

3.1.1. Fuzzy Sets

A Fuzzy Set A in a universe of discourse X is defined, according to FS theory [7], as a set of ordered pairs:

$$A = \{(x, \mu_A(x))/x \epsilon X\} \tag{1}$$

where the function $\mu_A : X \to [0,1]$ defines the degree of membership of the element $x \epsilon X$.

3.1.2. Intuitionistic Fuzzy Sets

Similarly, an Intuitionistic Fuzzy Set A in X is defined as follows [9]:

$$A = \{x, \mu_A(x)), \nu_A(x)/x \epsilon X\} \tag{2}$$

where the function $\nu_A : X \to [0,1]$ defines the degree of non-membership of an element $x \epsilon X$, such that

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1, \forall x \epsilon X \tag{3}$$

An important extension to IFS theory is the introduction of the hesitation (or hesitancy) degree of an element x belonging to a set A, defined as follows:

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{4}$$

It is obvious from Equations (3) and (4) that $0 \leq \pi_A(x) \leq 1, \forall x \epsilon X$. It should also be noted that in the case of FSs, an element's hesitation degree is $\pi_A(x) = 0, \forall x \epsilon X$, which can be described as a particular case of an IFS.

3.1.3. Distance Measure

A distance measure describes how different elements of a set are, and its definition is described as follows.

Definition 1. *A distance d in a nonempty set A is a real function $d : A \times A \rightarrow [0, +\infty)$ that satisfies the following conditions, $\forall x, y, z \epsilon A$ [20]:*

C1 $d(x, y) = 0 \Leftrightarrow x = y$ *(coincidence);*
C2 $d(x, y) = d(y, x)$ *(symmetry);*
C3 $d(x, z) + d(z, y) \geq d(x, y)$ *(triangle inequality).*

A very simple function to measure the distance between two FSs A and B is the Euclidean distance [21]:

$$e_1(A, B) = \sqrt{\sum_{i=1}^{n} (\mu_A(x_i) - \mu_B(x_i))^2} \tag{5}$$

The above Equation (5) was extended for application in IFSs to incorporate the non-membership values $\nu(.)$ [22]:

$$e_2(A, B) = \sqrt{\frac{1}{2} \sum_{i=1}^{N} (\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2} \tag{6}$$

as well as the incorporation of the hesitation degrees [23]:

$$e_3(A, B) = \sqrt{\frac{1}{2} \sum_{i=1}^{N} (\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2} \tag{7}$$

3.1.4. Similarity Measure

A similarity measure is defined as follows:

Definition 2. *A similarity S in a nonempty set A is a real function $S : A \times A \rightarrow [0, 1]$ that satisfies the following conditions, $\forall x, y, z \epsilon A$ [24–26]:*

C1 $S(x, y) = 1$ *if and only if* $x = y$ *(coincidence);*
C2 $S(x, y) = S(y, x)$ *(symmetry);*
C3 *if* $x \subseteq y \subseteq z$, *then* $S(x, z) \leq S(x, y)$ *and* $S(x, z) \leq S(y, z)$ *(triangle inequality).*

The first similarity measure was proposed by S. M. Chen [27]:

$$S(A, B) = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{|S_A(x_i) - S_B(x_i)|}{2}\right) \tag{8}$$

where $S_k(x_i) = \mu_k(x_i) - \nu_k(x_i), k = \{A, B\}$.

3.2. Evaluation Protocol

The protocol followed in our study was proposed by P. Intarapaiboon [10]. This protocol performs a pattern learning process, where the IFS patterns of each class of document are learned. Using those learned patterns, any new document can be classified by extracting its IFS patterns. The classification process is performed by finding the class in which the IFS patterns have the lowest distance or highest similarity with the new document.

Figure 1 shows the preprocessing step of the protocol, which includes the document preprocessing and the IFS pattern learning. In this step, the raw documents are cleaned, converted into a bag-of-words vector, and then, transformed into the IFS-based representation. Lastly, the IFS-based representation is used to calculate the IFS patterns of each class of document (called pattern learning).

Figure 1. Document learning protocol used in this study.

At the testing step, a test document d_i is converted into the IFS representation, and the class C_k with the closest pattern to the test document is selected as the document's class. The closest pattern is determined by using a distance or similarity measure.

3.2.1. Bag-of-Words Representation

This type of representation is the most widely used method for representing a text document as numbers in NLP. Based on this process, a text document is represented by the frequency of occurrence of each word. For a bag-of-words to be formulated, it requires a vocabulary of the known words (the words to be included) or the usage of a threshold value to cut off words with low frequency. Thus, a document's d_i bag-of-words vector V for h words can be represented as follows:

$$V = (n_{i,1}, n_{i,2}, n_{i,3}, \ldots, n_{i,h}) \tag{9}$$

where n is the number of occurrences of each word of the vocabulary.

3.2.2. IFS Representation

To transform a bag-of-words representation of a document d_i into an IFS-based representation, the following steps were followed (as proposed by [10]). Let us define

$$z_{i,j} = \frac{n_{i,j} - mean_j}{std_j} \tag{10}$$

where $n_{i,j}$ is the number of occurrences of word j in d_i and $mean_j$ and std_j are the mean and standard deviation of occurrences of word j in all documents of each class.

Thus, to calculate the membership and non-membership values of d_i, a weighted sigmoid function is used:

$$\mu_{i,j} = \frac{r_j}{1 + e^{-z_{i,j}}} \tag{11}$$

$$\nu_{i,j} = \frac{r_j^*}{1 + e^{z_{i,j}}} \tag{12}$$

where r_j and r_j^* are weights $\epsilon[0,1]$. The hesitation degree is then calculated using Equation (4).

3.2.3. IFS Pattern Learning

Lastly, for the pattern learning step, the IFS pattern of class C_k is denoted by P_k, is defined by:

$$P_k = \{(\tilde{\mu}_k, \tilde{\nu}_k)\} \tag{13}$$

where $\tilde{\mu}_k$ and $\tilde{\nu}_k$ are the average values of the membership and non-membership values of all the documents belonging to class C_k.

3.2.4. Document Classification

A document's class is determined according to how dissimilar (for distances) or similar (for similarities) its patterns are to the class patterns calculated in the previous step. To calculate a new document's IFS representation, the $mean_j$ and std_j values calculated during the class pattern learning step are used.

Therefore, to determine a new document's d_n class C', the following equations can be used to classify a test document:

$$C' = arg \min_{C_k} \{Dist(P_k, IFS_{d_n})\} \tag{14}$$

$$C' = arg \max_{C_k} \{Sim(P_k, IFS_{d_n})\} \tag{15}$$

where $Dist$ and Sim are the Distance and Similarity measures, respectively between the IFSs, P_k are the class patterns calculated during the IFS pattern learning step for class C_k, and IFS_{d_n} is the IFS representation of the new document d_n.

3.3. Experimental Setup

3.3.1. Distance and Similarity Measures

In this study, we evaluated the performance of 43 similarity measures and 19 distance measures, all provided by the fsmpy Python library [28]. Tables 1 and 2 show the names of the authors that proposed the corresponding measure in the first column, the parameters (other than the two Intuitionistic Fuzzy Sets A and B and the type of similarity) that are required by the functions to calculate the corresponding measure, and in the third column, the name of the proposed function that the authors used in the original study to mathematically describe it. As two of the measure d_h proposed by [22,29] have the same name, the latter one was renamed to d_{h2}. Furthermore, all the mathematical expressions of the measures are presented in Appendix A.

Table 1. Similarity measures evaluated in the study and provided by the fsmpy library.

Authors Names	Parameters	Measure Name	Reference
S. M. Chen	weights, a, b, c	S_e	[30]
D.H. Hong, C. Kim	weights, a, b, c	M_H	[31]
L. Dengfeng, C. Chuntian	p, weights	S_d^p	[24]
Z. Liang, P. Shi	p, weights	S_e^p, S_s^p	[32]
H.B. Mitchell	p, weights	S_{mod}	[25]
W.L. Hung, M.S. Yang	weights	S_c, S_e, S_l	[26]
H.W. Liu	p, weights, a, b, c	T	[33]
C. Zhang, H. Fu	-	S	[34]
W.L. Hung, M.S. Yang	a	S_p^c, S_p^e, S_p^l	[35]
S. Park, Y.C. Kwun, K.M. Lim	p, weights	S_{gw}^p	[36]
W.L. Hung, M.S. Yang	-	S_{new2}, S_{pk1}, S_{pk2}, S_{pk3}, S_{w1}, S_{w2}	[37]
W.L. Hung, M.S. Yang	p	S_a^c, S_a^e, S_a^l	[38]
J. Ye	weights	C_{IFS}	[39]
C.M. Hwang, M.S. Yang	-	S_{IFS}	[40]
P. Julian, K.C. Hung, S. Lin	p, weights	$S_{new,p}$	[41]
I. Iancu	lamda	$\tilde{S}_1$, $\tilde{S}_5$, $\tilde{S}_2$, $\tilde{S}_6$, $\tilde{S}_2'$, $\tilde{S}_6'$, $\tilde{S}_1^c$, $\tilde{S}_5^c$	[42]
G. Deng, Y. Jiang, J. Fu	weights, p, u, v	L_5, N_6, F_5, F_6, F_7	[43]
Y. Song, X. Wang, L. Lei, A. Xue	weights	S_{WY}	[44]
S.M. Chen, S. Cheng, T.C. Lan	weights	S_{CCL}	[45]
P. Muthukumar, G.S.S. Krishnan	weights	W_{IFSSs}	[46]

Table 2. Distance measures evaluated in the study and provided by the fsmpy library.

Authors Names	Parameters	Measure Name	Reference
K.T. Atanassov	-	d_e, d_h, d_{ne}, d_{nh}	[22]
E. Szmidt, A. Kacprzyk	-	d^1_{IFS}, e^1_{IFS}, l^1_{IFS}, q^1_{IFS}	[23]
P. Grzegorzewski	-	d_{h2}, e_h, l_h, q_h	[29]
W. Wang, X. Xin	weights, p	d_1, d_2^p	[20]
Y. Yang, F. Chiclana	-	d_{eh}, l_{eh}, e_{eh}, q_{eh}	[47]
I.K. Vlachos, G.D. Sergiadis	-	D_{IFS}	[48]

At this point, it should be mentioned that not all of the measures available in the fsmpy library were compared in this study. A total of 27 measures were not included as their implementations are not trustworthy enough, failing to pass the required tests by the library. Specifically, the rest of the measures proposed by [32] (S_h^p), [43] (namely, L_3, L_4, L_6, N_4, N_5, and N_7), [37] (S_{new1}) [42] (namely, $\tilde{S}_{11}$, $\tilde{S}_{19}$, $\tilde{S}_{14}$, $\tilde{S}_{18}$, $\tilde{S}_{13}$, $\tilde{S}_{15}$, $\tilde{S}'_{13}$, $\tilde{S}'_{15}$, $\tilde{S}^c_{11}$, $\tilde{S}^c_{19}$, $\tilde{S}^c_{14}$, $\tilde{S}^c_{18}$, $\tilde{S}^c_{2}$, $\tilde{S}^c_{6}$, $\tilde{S}^c_{13}$, $\tilde{S}^c_{15}$), [49] (S_F) and [27] (T) were not included.

It is worth mentioning that in order to optimize the performance of the measures, a hyperparameter optimization technique was employed using both the fsmpy and Scikit-Optimize libraries. For the optimization process, we employed the Bayesian optimization technique, which, contrary to the Grid Search optimization technique, does not try out all the parameter values, but only a fixed number of parameter settings from specified distributions. The performance of each parameter combination was tested over a 10-fold cross-validation technique, with the best one being selected according to the highest F1-score metric. It should be noted that the weights parameter was set to $weights = \overrightarrow{1/n}$, where n is the number of elements in the set.

3.3.2. Datasets

The datasets that were used in our experiments are presented and described below:

1. BBC News: This consists of 2225 articles belonging to 5 topic areas (business, entertainment, politics, sports, and tech) from 2004 and 2005, with a total of 9635 words [50].
2. BBC Sports: Similar to the previous one, it contains 737 articles from 5 areas, namely athletics, cricket, football, rugby, and tennis, having a total of 4613 words [50].

Figure 2 shows, for each dataset, the number of samples for each class, along with the mean number of words of the sample texts. More specifically, as shown in Figure 2a, the classes in the BBC News dataset have a balanced distribution of samples, with 386 entertainment-type and 511 sports-type documents. The classes have a balanced word count as well, with the texts belonging to the class business having a mean word count of 339 (the lowest) and tech having 513 (the highest).

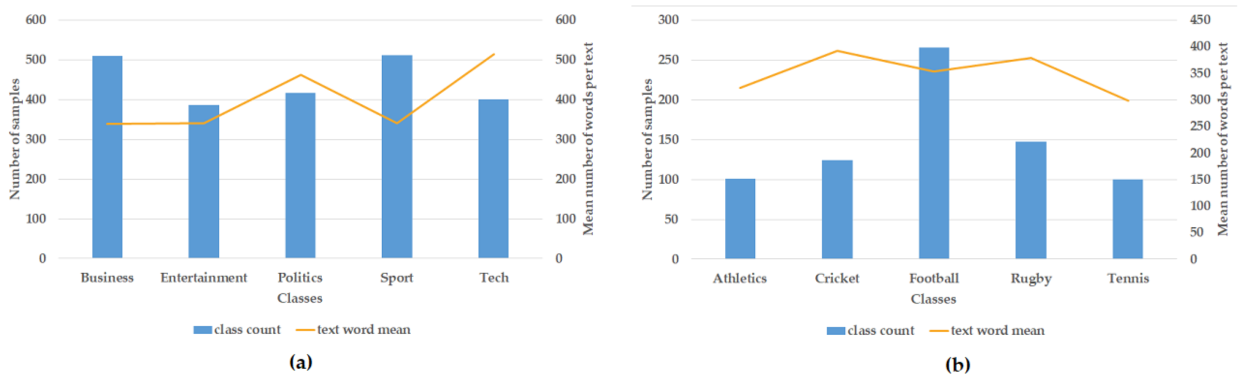

Figure 2. Class distribution and mean of word count per sample text for (**a**) BBC News and (**b**) BBC Sports datasets.

For the BBC Sports case though, the distribution of the samples is not as balanced. As shown in Figure 2b, the football class has the highest number of samples (265), with the rest of the classes having 100–150 samples On the other hand, the mean number of words per text is balanced across all classes, with the tennis class having a mean word count of 298 (lowest) and cricket having 391 (highest).

3.3.3. Data Preparation

The data preparation step consisted of the following steps:

1. Documents' conversion to bag-of-words;
2. Word-by-word preprocessing:
   (a) Remove escape characters;
   (b) Convert to lower case;
   (c) Lemmatize word.
3. Cut-off words with a document frequency lower than a 0.01 frequency.

It is worth noting that lemmatization is the process of converting variations of a word to its base form, for example: "does", "doing", and "did" were converted to "do" and "am" and "are" and "is" to "be". For this process, the Scikit-learn [51] and Natural Language Toolkit [52] libraries were used. Moreover, the chosen by trial and error cut-off frequency resulted in the BBC News dataset containing a bag-of-words of 2540 words and BBC Sports 2396 words.

3.3.4. Performance Evaluation

The performance of each measure was evaluated using the accuracy, precision, recall, and F1-score metrics, as well as the Degree of Confidence (DoC) proposed by A.G. Hatzimichailidis et al. [53]. To calculate the Degree of Confidence metric, the fsmpy library was used and the Scikit-learn library for the rest of the metrics. These metrics are widely used in classification studies to evaluate the performance of an algorithm or model, with the DoC being used in the literature on FS theory.

More specifically, the DoC measures the confidence at which the measure can recognize a specific sample belonging to a specific pattern. As a result, higher values of the DoC show that the measure has more confidence when predicting a sample's class. Moreover, to calculate the precision, recall, and F1-score, we used their "macro" average, calculating the metrics for each label and finding their unweighted mean.

4. Results

In our study, we evaluated the performance of 43 similarity measures and 19 distance measures, in the classification of text documents of two different datasets. The experiments were conducted using the Python programming language along with the Scikit-learn and fsmpy libraries. Moreover, we examined the performance of each parameter group by following a grid search technique for r_j and r_j^*, by selecting five linearly spaced values over the interval $[0.1, 1]$. Specifically, the r_j and r_j^* values that were tested were the following: $\{0.1, 0.325, 0.55, 0.775, 1\}$

After finding the best parameters of each measure, along with their best membership and non-membership weights, we validated their performance on each dataset by employing a leave-one-out cross-validation technique. In this validation technique, a single sample is considered for testing in each fold.

Tables 3–6 show the performance of the similarity and distance measures on each dataset. Each table shows the name of the measure as mentioned in the corresponding study that introduced it, the best parameters (where applied) obtained from the hyperparameter optimization, the best membership and non-membership weights obtained, the accuracy, precision, recall, and F1-score performance metrics, and the mean DoC of the measure.

Specifically, Tables 3 and 4 show the performance of the similarity and distance measures on the BBC News dataset and Tables 5 and 6 their performance on the BBC Sports dataset.

Table 3. Performance of distance measures on BBC News dataset.

Distance Measures	Parameters	Weights $(r_{j_1}, r^*_{j_1})$	Accuracy %	Precision %	Recall %	F1 %	DoC
d_e	-	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0903
d_h	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	3.9711
d_{ne}	-	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
d_{nh}	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
d^1_{IFS}	-	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0903
e^1_{IFS}	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	3.9711
l^1_{IFS}	-	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
q^1_{IFS}	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
d_h	$p = 7.00$	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
e_h	$p = 7.00$	(0.1, 0.1)	88.94	90.45	88.98	88.63	0.0031
l_h	-	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0903
q_h	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	3.9711
d_1	-	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
d^p_2	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
d_{eh}	-	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0903
l_{eh}	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	3.9711
e_{eh}	-	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
q_{eh}	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
D_{IFS}	-	(0.1, 1.0)	**94.25**	**94.10**	**93.98**	**94.02**	**5.5846**

Table 4. Performance of similarity measure on BBC News dataset.

Similarity Measures	Parameters	Weights $(r_{j_1}, r^*_{j_1})$	Accuracy %	Precision %	Recall %	F1 %	DoC
S_e	$a = 6.00$, $b = -4.00$, $c = 7.00$	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
M_H	$a = 6.00$, $b = -4.00$, $c = 7.00$	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
S^p_d	$p = 7.00$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
S^p_e	$p = 7.00$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
S^p_s	$p = 7.00$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
S_{mod}	$p = 7.00$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
S_c	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
S_e	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0025
S_l	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0031
T	$a = 0.34$, $b = 0.47$, $c = 0.42$, $p = 9.00$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018

Table 4. *Cont.*

Similarity Measures	Parameters	Weights $(r_{j_1}, r^*_{j_1})$	Accuracy %	Precision %	Recall %	F1 %	DoC
S	-	(0.1, 1.0)	50.65	61.21	49.05	39.49	0.0091
S^c_p	$a = 6.00$	(1.0, 0.1)	**94.83**	**94.67**	**94.62**	**94.63**	0.0019
S^e_p	$a = 6.00$	(1.0, 0.1)	**94.83**	**94.67**	**94.62**	**94.63**	0.0021
S^l_p	$a = 6.00$	(1.0, 0.1)	**94.83**	**94.67**	**94.62**	**94.63**	0.0022
S^p_{gw}	$p = 7.00$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0025
S_{new2}	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0023
S_{pk1}	-	(1.0, 0.1)	50.47	60.86	48.9	39.26	0.0293
S_{pk2}	-	(1.0, 0.1)	75.06	77.75	74.99	74.56	**0.0418**
S_{pk3}	-	(1.0, 0.1)	50.47	60.86	48.9	39.26	0.0157
S_{w1}	-	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0017
S_{w2}	-	(0.1, 0.1)	49.17	39.46	47.68	37.84	0
S^c_a	$p = 7.00$	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0026
S^e_a	$p = 7.00$	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0033
S^l_a	$a = 6.00$, $b = -4.00$, $c = 7.00$	(0.1, 0.1)	49.66	61.04	48.18	38.51	0.0016
C_{IFS}	-	(0.1, 0.325)	94.74	94.67	94.57	94.56	0.0033
S_{IFS}	-	(0.55, 0.325)	93.71	93.44	93.47	93.44	0.0019
$S_{new,p}$	$p = 7.00$	(0.325, 0.1)	93.17	93.19	92.76	92.86	0.0040
$\tilde{S}_1$	-	(0.1, 1.0)	71.91	80.40	73.01	73.03	0.0097
$\tilde{S}_5$	-	(1.0, 0.1)	59.78	75.68	60.60	61.30	0.0112
$\tilde{S}_2$	-	(0.55, 0.1)	75.55	84.69	74.10	73.14	0.0007
$\tilde{S}_6$	-	(0.1, 0.55)	91.69	92.20	91.84	91.49	0.0006
$\tilde{S}'_2$	-	(1.0, 1.0)	55.33	58.77	56.51	55.42	0.0091
$\tilde{S}'_6$	-	(0.775, 0.325)	93.48	93.36	93.45	93.29	0.0012
$\tilde{S}^c_1$	-	(1.0, 0.325)	64.9	76.35	65.71	66.39	0.0056
$\tilde{S}^c_5$	-	(0.1, 0.1)	90.25	91.11	89.72	89.72	0
L_5	$p = 7.00$, $u = 4.43$, $v = 7.22$	(0.775, 0.55)	94.29	94.24	94.09	94.10	0.0289
N_6	$p = 7.00$, $u = 4.43$, $v = 7.22$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0018
F_5	$p = 7.00$, $u = 4.43$, $v = 7.22$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0006
F_6	$p = 7.00$, $u = 4.43$, $v = 7.22$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0
F_7	$p = 7.00$, $u = 4.43$, $v = 7.22$	(0.1, 0.1)	93.17	93.19	92.76	92.86	0.0028
S_{WY}	-	(0.1, 0.325)	94.47	94.26	94.23	94.24	0.0005
S_{CCL}	-	(0.1, 1.0)	49.93	60.98	48.42	38.74	0.0083
W_{IFSSs}	-	(1.0, 0.1)	65.75	81.88	63.86	59.92	0.0285

Table 5. Performance of similarity measure on BBC Sports dataset.

Similarity Measures	Parameters	Weights $(r_{j_1}, r^*_{j_1})$	Accuracy %	Precision %	Recall %	F1 %	DoC
S_e	$a = 6.00$, $b = -4.00$, $c = 7.00$	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.0014
M_H	$a = 6.00$, $b = -4.00$, $c = 7.00$	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.0014
S_d^p	$p = 7.00$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0028
S_e^p	$p = 7.00$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0028
S_s^p	$p = 7.00$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0028
S_{mod}	$p = 7.00$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0028
S_c	-	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.0014
S_e	-	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.0023
S_l	-	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.0028
T	$a = 0.34$, $b = 0.47$, $c = 0.42$, $p = 9.00$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0028
S	-	(0.1, 1.0)	28.49	83.19	36.36	29.31	0.0081
S_p^c	$a = 6.00$	(0.1, 0.1)	96.47	96.69	**96.78**	96.72	0
S_p^e	$a = 6.00$	(0.1, 0.1)	96.47	96.69	**96.78**	96.72	0
S_p^l	$a = 6.00$	(0.1, 0.1)	96.47	96.69	**96.78**	96.72	0
S_{gw}^p	$p = 7.00$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0045
S_{new2}	-	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.0021
S_{pk1}	-	(0.775, 0.1)	26.46	83.12	35.38	27.44	0.0269
S_{pk2}	-	(0.775, 0.325)	79.65	81.51	77.73	78.79	**0.0489**
S_{pk3}	-	(0.775, 0.1)	26.46	83.12	35.38	27.44	0.0144
S_{w1}	-	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.002
S_{w2}	-	(0.1, 0.1)	24.42	81.04	33.44	24.17	0
S_a^c	$p = 7.00$	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.0024
S_a^e	$p = 7.00$	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.003
S_a^l	$a = 6.00$, $b = -4.00$, $c = 7.00$	(0.1, 0.1)	25.10	83.07	33.90	25.50	0.0014
C_{IFS}	-	(0.1, 0.55)	**96.47**	**96.99**	96.71	**96.84**	0.0023
S_{IFS}	-	(1.0, 0.775)	95.93	96.54	96.11	96.31	0.0067
$S_{new,p}$	$p = 7.00$	(0.325, 0.1)	96.34	96.54	96.47	96.48	0.0064
$\tilde{S}_1$	-	(0.1, 1.0)	69.61	76.86	72.68	71.01	0.0072
$\tilde{S}_5$	-	(1.0, 0.1)	50.47	78.02	54.83	54.77	0.0096
$\tilde{S}_2$	-	(0.775, 0.1)	82.36	86.95	87.83	84.91	0.0012
$\tilde{S}_6$	-	(0.1, 0.55)	94.44	93.89	95.63	94.61	0.0006
$\tilde{S}_2'$	-	(1.0, 1.0)	40.30	58.81	47.36	43.41	0.0072
$\tilde{S}_6'$	-	(1.0, 0.1)	94.30	93.71	95.61	94.50	0.0018
$\tilde{S}_1^c$	-	(1.0, 0.325)	63.50	73.42	66.40	65.26	0.004
$\tilde{S}_5^c$	-	(0.1, 0.1)	90.09	89.41	91.95	89.84	0
L_5	$p = 7.00$, $u = 4.43$, $v = 7.22$	(1.0, 0.775)	96.20	96.46	96.34	96.37	0.0371

Table 5. *Cont.*

Similarity Measures	Parameters	Weights $(r_{j_1}, r^*_{j_1})$	Accuracy %	Precision %	Recall %	F1 %	DoC
N_6	$p = 7.00$, $u = 4.43$, $v = 7.22$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0028
F_5	$p = 7.00$, $u = 4.43$, $v = 7.22$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0009
F_6	$p = 7.00$, $u = 4.43$, $v = 7.22$	(1.0, 0.775)	96.20	96.36	96.40	96.35	0
F_7	$p = 7.00$, $u = 4.43$, $v = 7.22$	(0.1, 0.1)	96.34	96.54	96.47	96.48	0.0045
S_{WY}	-	(0.1, 0.325)	95.93	97.33	95.64	96.42	0.0006
S_{CCL}	-	(0.1, 1.0)	25.78	83.09	34.40	26.22	0.0075
W_{IFSSs}	-	(1.0, 0.1)	62.82	84.04	65.53	64.82	0.0261

Table 6. Performance of distance measures on BBC Sports dataset.

Distance Measures	Parameters	Weights $(r_{j_1}, r^*_{j_1})$	Accuracy %	Precision %	Recall %	F1 %	DoC
d_e	-	(0.1, 0.1)	93.62	96.32	92.79	94.28	0.1001
d_h	-	(0.1, 0.1)	25.1	83.07	33.9	25.5	3.423
d_{ne}	-	(0.1, 0.1)	93.62	96.32	92.79	94.28	0.002
d_{nh}	-	(0.1, 0.1)	25.1	83.07	33.9	25.5	0.0014
d^1_{IFS}	-	(0.1, 0.1)	93.62	96.32	92.79	94.28	0.1001
e^1_{IFS}	-	(0.1, 0.1)	25.1	83.07	33.9	25.5	3.423
l^1_{IFS}	-	(0.1, 0.1)	93.62	96.32	92.79	94.28	0.002
q^1_{IFS}	-	(0.1, 0.1)	25.1	83.07	33.9	25.5	0.0014
d_h	$p = 7.00$	(0.1, 0.1)	25.1	83.07	33.9	25.5	0.0014
e_h	$p = 7.00$	(0.1, 0.1)	95.66	95.54	96.15	95.77	0.0034
l_h	-	(0.1, 0.1)	93.62	96.32	92.79	94.28	0.1001
q_h	-	(0.1, 0.1)	25.1	83.07	33.9	25.5	3.423
d_1	-	(0.1, 0.1)	93.62	96.32	92.79	94.28	0.002
d^p_2	-	(0.1, 0.1)	25.1	83.07	33.9	25.5	0.0014
d_{eh}	-	(0.1, 0.1)	93.62	96.32	92.79	94.28	0.1001
l_{eh}	-	(0.1, 0.1)	25.1	83.07	33.9	25.5	3.423
e_{eh}	-	(0.1, 0.1)	93.62	96.32	92.79	94.28	0.002
q_{eh}	-	(0.1, 0.1)	25.1	83.07	33.9	25.5	0.0014
D_{IFS}	-	(0.1, 0.55)	**95.39**	**97.07**	**95.02**	**95.94**	**3.6342**

5. Discussion

In general, the application of the evaluation protocol proposed by [10] shows very good classification performance, in most of the similarity or distance measures. First of all, the most important aspect of such an approach for text classification is that it requires no training process. As a result, the performance is not affected by any imbalanced class distribution, with the presented study highlighting this advantage. As shown in Figure 2, the class distribution in the BBC Sports dataset is imbalanced, which would result in an overfitting towards one of the classes if a machine learning model were applied.

Secondly, the results highlight the importance of which IFS measure is used for this type of classification problem. Although most of the measures performed above 90% in accuracy and F1-score, there were a few that performed very poorly.

Thirdly, similarity measures showed very poor prediction confidence, with the distance measure (on the contrary) showing very confident predictions. Despite that, they performed (on average) better than the distance measures, although with a very small difference. Thus, the experimental results showed that the distance should be preferred, as it had high classification performance with very good confidence.

Lastly, by comparing the types of similarities, it can be observed that there was no change in the performance, showing that there was no significant difference between them. This can be highlighted more when taking into account the best parameters obtained from the hyperparameter optimization process, where measures of the same type had the same best parameters, as well as membership and non-membership weights.

Below, we discuss the results obtained for each dataset in more detail.

5.1. BBC News Results

It is obvious that the results obtained from different similarity measures, as well as distance measures were quite similar. Starting with the results obtained from the similarity measures, the ones that performed the best were from W.L. Hung and M.S. Yang [35] with 94.83% accuracy, 94.67% precision, 94.62% recall, and 94.63% F1-score. The next-best-performing measure was by J. Ye [39] with 91.70% and Y. Song, X. Wang, L. Lei, and A. Xue [44] with a 94.56% F1-score. Despite the best performance regarding those metrics, their DoC was not the highest as S_{pk2} had the highest DoC despite having much lower performance. There were, though, quite a few measures that performed below 50%, with the worst performance being at 49.17% accuracy, 39.46% precision, 47.68% recall, 37.84% F1-score, and 0 DoC by S_{w2}. Therefore, these types of measures are not well suited for document classification, as they fail to capture those important similarities between the Intuitionistic Fuzzy Sets of the documents and classes. Moreover, the measures' performance regarding the DoC showed that there were no large differences between the similarities calculated for the chosen classes compared to the other classes. This shows that, despite the very good performance of some measures (or even all of them), they were not that confident for their prediction.

As for the distance measures, the best-performing one was that of I.K. Vlachos and G.D. Sergiadis [48] with 94.25% accuracy, 94.10% precision, 93.98% recall, and 94.02% F1-score, followed by all the Euclidean-based distances, specifically d_e, d_{ne}, d^1_{IFS}, l^1_{IFS}, l_h, d_1, d_{eh}, and e_{eh} with 93.17% accuracy, 93.19% precision, 92.76% recall, and 92.86% F1-score. In contrast with the DoC of the similarities, the distances seemed to be more confident about their prediction, with D_{IFS} having the highest confidence of 5.5846, showing large distances between the different classes.

5.2. BBC Sports Results

In the case of the BBC Sports dataset, similar ranking results were obtained, but with better performance. Regarding the similarity measures, the measure proposed by J. Ye [39] performed the best with 96.47% accuracy, 96.99% precision, 96.71% recall, 96.84% F1-score, and 0.0023 DoC, followed by the measures proposed by W.L. Hung and M.S. Yang [35] with 96.47% accuracy, 96.69% precision, 96.78% recall, and 96.72% F1-score. The most confident measure was again S_{pk2}, despite performing much worse than C_{IFS}.

As for the distance measures, the one proposed by I.K. Vlachos and G.D. Sergiadis [48] performed the best, with 95.39% accuracy, 97.07% precision, 95.02% recall, 95.94% F1-score, and 3.6342 DoC, followed by the same Euclidean-based measures, as in the previous dataset, with 93.62% accuracy, 96.32% precision, 92.69% recall, and 94.28% F1-score. Similarly, the distances showed higher confidence about their predictions, highlighting that the distances between the classes were larger.

5.3. Comparison with Standard Machine Learning Approaches

To better assess the application of the evaluation protocol, as well as the IFS measures included in the study, we compared the best results obtained for each case with standard ML models on the same datasets. For the comparison, we employed the Decision Tree (DT) and K-Nearest Neighbors (KNN) ML models. Both models were trained on the same datasets, by following the same data preparation, performance evaluation, and hyperparameter optimization processes. Table 7 shows the results obtained from the evaluation of the aforementioned models, along with the results from the best performing similarity and distance measures in both datasets.

Table 7. Performance of the ML models and the best-performing IFS measures on both datasets.

	BBC News				BBC Sports			
Model/ Measure	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 (%)**
Decision Tree	81.61	81.85	81.04	81.31	88.19	88.14	87.87	87.96
KNN	82.69	86.47	81.63	82.79	88.46	92.96	87.42	89.16
S_p^c	**94.83**	**94.67**	**94.62**	**94.63**	96.47	96.69	**96.78**	96.72
C_{IFS}	94.74	94.67	94.57	94.56	**96.47**	**96.99**	96.71	**96.84**
D_{IFS}	94.25	94.10	93.98	94.02	95.39	97.07	95.02	95.94

The results obtained from the conventional ML models highlight the advantages of the IFS-based classification approaches and IFS measures in general. The first advantage is the better performance of the IFS-based approach on both datasets, showing an improvement of 12% on the BBC News dataset and 10% on the BBC Sports dataset. The best-performing ML model, according to the F1-score metric, on both datasets was the KNN classifier with an 82.79% and 89.16% F1-score, respectively. The Decision Tree model performed at about the same, with 81.61% accuracy, 81.85% precision, 81.04% recall, and 81.31% F1-score, a small difference compared to the KNN.

The second advantage of the IFS-based text classification is that it does not include any training process, by-passing the issue of class imbalance and overfitting on the class with the most samples. This is a very important characteristic, as datasets tend to be very imbalanced, with many methodologies being proposed to tackle this issue. The overfitting of the model (and lack thereof on the S_p^c) can be observed in Figure 3, where the confusion matrices are shown during the validation of both. In the first case of the DT (Figure 3a), the model overfitted towards the football class, falsely predicting samples of other classes. On the other hand, the S_p^c (Figure 3b) similarity (or the evaluation protocol in general) was not affected by the class imbalance.

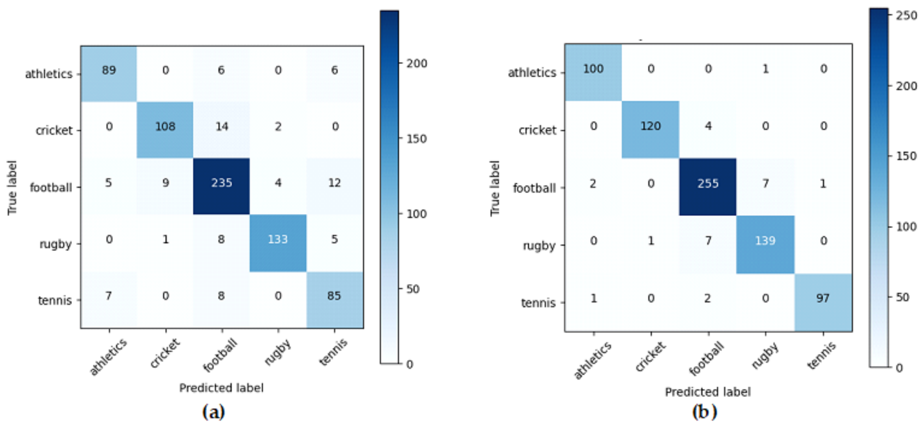

Figure 3. Confusion matrices of (**a**) Decision Tree and (**b**) S_p^c during the validation process.

6. Conclusions

In this study, we evaluated the performance of 62 Intuitionistic Fuzzy Set measures on the same base towards classifying texts from two datasets, for the first time in the literature. Their performance was evaluated by following the protocol proposed by P. Intarapaiboon [10], where the documents are represented as Intuitionistic Fuzzy Sets by firstly converting the texts into bag-of-words representations. Then, by using a sigmoid function to calculate the membership and non-membership values of the documents (and the classes), given the membership and non-membership weights, the IFS representation and patterns were obtained. Moreover, we optimized the performance of each measure separately, by finding the combination of parameter values and membership and non-membership weights that yielded the best results in a 10-fold cross-validation technique. Their performance was then validated by employing the leave-one-out cross-validation technique.

The results obtained highlight a number of important findings. Firstly, although the functions are different, in the case of text document classification, their performance was not significantly different, as the results obtained for the different measures were the same in most cases. This can be attributed to the fact that some studies aim to overcome the drawbacks of previous measures, such as the measure proposed by H.B. Mitchell [25], who revised the measure proposed by L. Dengfeng and C. Chuntian [24]. Secondly, Hamming-based distance measures had an F1-score lower than 50%, which shows that the problem of text classification is not as simple as calculating the absolute difference between two sets. Thirdly, the similarities maintained a very low degree of confidence during their predictions, which shows that the resulting class patterns were very similar. Moreover, the DoC requires similarities to be converted to distances via the formula $1 - S$. As a result, in combination with the definition of the similarities where $S : A \times A \to [0, 1]$, it is obvious that their confidence will be much lower compared to distances, whose definition "allows" them to have much higher values $d : A \times A \to [0, +\infty)$. Lastly, different parameters had different effects on each measure, for example in the case of S_p^e, for different values of the parameter a, the performance remained almost the same during the hyperparameter optimization technique, with an std of 0.7%, while L_5 had an std of 9% for different values of p.

Despite the wide range of different similarity measures included in the study, there were some measures that were not included. For instance, the similarity measure proposed by Z. Xu [54], which was also included in [10], was not evaluated in the study. Additionally, the types of measures evaluated in our study (distances and similarities) are the first ones that have been proposed and employed for problems such as pattern recognition, medical diagnosis, image segmentation, or classification in general. For this reason, our analysis and evaluation were focused on those. In the future, some other types of measures have to be analyzed in the same manner, such as divergence, correlation, inclusion, and others.

Author Contributions: Conceptualization, G.K.S. and G.A.P.; methodology, G.K.S., N.D. and K.D.A.; software, G.K.S., N.D. and K.D.A.; validation, G.K.S., N.D. and K.D.A.; investigation, G.K.S., N.D. and K.D.A.; resources, G.K.S., N.D. and K.D.A.; data curation, G.K.S., N.D. and K.D.A.; writing—original draft preparation, G.K.S. and K.D.A.; writing—review and editing, G.K.S. and G.A.P.; visualization, G.K.S.; supervision, G.A.P.; project administration, G.A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported by the MPhil program "Advanced Technologies in Informatics and Computers", hosted by the Department of Computer Science, International Hellenic University, Greece.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Tables A1–A3 show the mathematical expressions of the similarity measures and Table A4 all the distance measures presented in our study. To simplify the expressions below, we use the following notations:

$$\mu = \mu_A(i), \nu = \nu_A(i), \pi = \pi_A(i),$$
$$m = \mu_B(i), n = \nu_B(i), p = \pi_B(i),$$
$$\Delta_\mu = \mu - m, \Delta_\nu = \nu - n, \Delta_\pi = \pi - p$$
$$\sum = \sum_{i=1}^{N} \tag{A1}$$

with $i \epsilon \{1, 2, \ldots, N\}$ being the nth membership, non-membership, or hesitation degree of the set and w_i being the weight of $x_i \epsilon X = \{x_1, x_2, \ldots, x_N\}$, with $0 \le w_i \le 1$.

Table A1. Mathematical expressions of the reviewed similarity measures I.

Similarity Measure	Expression
S_e	$\sum w_i \left(1 - \left\vert \frac{(\mu - \nu) - (m - n)}{2} \right\vert \right) / \sum w_i$
M_H	$\sum w_i \left(1 - \frac{\vert a * \Delta_\mu + b * \Delta_\nu - c * (\Delta_\mu + \Delta_\nu) \vert}{a - b}\right) / \sum w_i$ where $a, b, c \ge 0$
S_d^p	$1 - \frac{1}{\sqrt[\rho]{N}} \sqrt[\rho]{\sum \vert(\mu + 1 - \nu) - (m + 1 - n)\vert^\rho}$ where $0 \le \rho \le +\infty$
S_e^p	$1 - \frac{1}{\sqrt[\rho]{N}} \sqrt[\rho]{\sum \left(\frac{\vert\Delta_\mu\vert + \vert\Delta_\nu\vert}{2}\right)^\rho}$
S_s^p	$1 - \frac{1}{\sqrt[\rho]{N}} \sqrt[\rho]{\sum (\phi_{s1}(i) + \phi_{s2}(i))^\rho}$ where $\phi_{s1}(i) = \vert m_{A1}(i) - m_{B1}(i)\vert/2$, $\phi_{s2}(i) = \vert m_{A2}(i) - m_{B2}(i)\vert/2$, $m_{k1}(i) = (\mu_k(i) + m_k(i))/2$, $m_{k2}(i) = (m_k(i) + 1 - \nu_k(i))/2$, $m_k(i) = (\mu_k(i) + 1 - \nu_k(i))/2$
S_{mod}	$\frac{1}{2}(S_d^p(\mu, m) + S_d^p(1 - \nu, 1 - n))$
S_c	$\frac{1 - d_H(A, B)}{1 + d_H(A, B)}$
S_e	$\frac{e^{-d_h(A,B)} - e^{-1}}{1 - e^{-1}}$
S_l	$1 - d_H(A, B)$
T	$1 - \sqrt[\rho]{\sum w_i [a * \vert\Delta_\mu\vert^\rho + b * \vert\Delta_\nu\vert^\rho + c * \vert\Delta_\pi\vert^\rho]}$ where $1 < p < +\infty$, $a, b, c \epsilon [0, 1]$ and $a + b + c = 1$
S	$1 - \frac{1}{2N} \sum (\vert\delta_A - \delta_B\vert + \vert\alpha_A - \alpha_B\vert)$ where $\delta_k = \mu_k + (1 - \mu_k - \nu_k)\mu_k$, $\alpha_k = \nu_k + (1 - \mu_k - \nu_k)\nu_k$, with $k = \{A, B\}$
S_p^c	$\frac{\sqrt[\rho]{2} - L_p(A, B)}{\sqrt[\rho]{2}(1 + L_p(A, B))}$ where $L_p = \frac{1}{N} \sum (\vert\Delta_\mu\vert^\rho + \vert\Delta_\nu\vert)^\rho$
S_p^e	$\frac{e^{-L_p(A,B)} - e^{-\sqrt[\rho]{2}}}{1 - e^{-\sqrt[\rho]{2}}}$
S_p^l	$\frac{\sqrt[\rho]{2} - L_p(A, B)}{\sqrt[\rho]{2}}$

Table A2. Mathematical expressions of the reviewed similarity measures II.

Similarity Measure	Expression
S_{gw}^{p}	$1-\sqrt[\rho]{\sum w_i\left(\frac{\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert+\lvert\Delta_\pi\rvert}{2}\right)^{\rho}}$ where $0\le\rho\le+\infty$
S_{new2}	$1-\frac{1-\exp\left(1-\frac{1}{2}\sum(\lvert\sqrt{\mu}-\sqrt{m}\rvert+\lvert\sqrt{\nu}-\sqrt{n}\rvert)\right)}{1-e^{-n}}$
S_{pk1}	$\frac{\sum min(\mu,m)+min(\nu,n)}{\sum max(\mu,m)+max(\nu,n)}$
S_{pk2}	$1-\frac{1}{2}\left(max_i(\lvert\Delta_\mu\rvert)+max_i(\lvert\Delta_\nu\rvert)\right)$
S_{pk3}	$1-\frac{\sum(\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert)}{\sum(\lvert\mu+m\rvert+\lvert\nu,n\rvert)}$
S_{w1}	$\frac{1}{N}\frac{\sum min(\mu,m)+min(\nu,n)}{\sum max(\mu,m)+max(\nu,n)}$
S_{w2}	$\frac{1}{N}\sum\left(1-\frac{1}{N}(\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert)\right)$
S_a^c	$\frac{U(a)-J_a(A,B)}{(1+J_a(A,B))U(a)}$ where $U(a)=\begin{cases}\ln 2 & a=1\\ \frac{1}{a-1}\left(1-\frac{1}{2^{a-1}}\right) & a\neq 1, a>0'\end{cases}$ $J_a(A,B)=\frac{1}{N}\sum j_a(A_i,B_i)$, $j_a(A,B)=\begin{cases}\frac{-1}{a-1}T_{AB}^{\mu a}T_{AB}^{\nu a}T_{AB}^{\pi a} & a\neq 1, a>0\\ \frac{-1}{2}L_{AB}^{\mu}L_{AB}^{\nu}L_{AB}^{\pi} & a=1\end{cases}$ $T_{AB}^{qa}=\left(\frac{q_A+q_B}{2}\right)^{a}-\frac{1}{2}(q_A^a+q_B^a)$ $L_{AB}^{q}=(q_A+q_B)ln\left(\frac{q_A+q_B}{2}\right)-q_A\ln q_A-q_B\ln q_B$, with $q=\{\mu,\nu,\pi\}$
S_a^e	$\frac{e^{-J_a(A,B)}-e^{-U(a)}}{1-e^{-U(a)}}$
S_a^l	$\frac{U(a)-J_a(A,B)}{U(a)}$
C_{IFS}	$\frac{1}{N}\sum\frac{+p}{\sqrt{\mu_A^2(i)+\nu_A^2(i)}\sqrt{\mu_B^2(i)+\nu_B^2(i)}}$
S_{IFS}	$\frac{1}{3}(C_{IFS}+C_{IFS}^{*}+C_{IFS}^{**})$ where $C_{IFS}^{*}=\frac{1}{N}\sum\frac{\phi_A\phi_B+\nu n}{\sqrt{\phi_A^2\nu^2}\sqrt{\phi_B^2 n^2}}$, $C_{IFS}^{**}=\frac{1}{N}\sum\frac{(1-\mu)(1-m)+(1-\nu)(1-n)}{\sqrt{(1-\mu)^2+(1-\nu)^2}\sqrt{(1-m)^2+(1-n)^2}}$ with $\phi_k=\frac{1+\mu_k-\nu_k}{2}$
$S_{new,p}$	$1-\sqrt[\rho]{\sum w_i(\lvert\Delta_\mu\rvert)^{\rho}}-\sqrt[\rho]{\sum w_i(\lvert\Delta_\nu\rvert)^{\rho}}$ where $w_i 0$, $\rho 1$
$\tilde{S}_1$	$\frac{N+\sum(T(\mu,m)+T(\nu,n)-\nu_A-\nu_B}{N+max\{\sum(\mu-m),\sum(\nu-p)\}}$ $T_\lambda(x,y)=\begin{cases}min(x,y) & if \lambda=0\\ x*y & if\lambda=1\\ max(0,x+y-1) & if\lambda=\infty\\ \log_\lambda(1+\frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}) & otherwise\end{cases}$
$\tilde{S}_5$	$\frac{N+\sum(T(\mu,m)+T(\nu,n)-\nu_A-\nu_B)}{n-\mu+m-T(\mu,m)-T(\nu,n)\}}$
$\tilde{S}_2$	$\frac{2N+\sum(2T(\mu,m)+2T(\nu,n)-\mu-m-\nu-n)}{2N+\sum(T(\mu,m)+T(\nu,n))-min\{\sum(\mu+p,\sum(m+\nu)\}}$
$\tilde{S}_6$	$\frac{2N+\sum(2T(\mu,m)+2T(\nu,n)-\mu-m-\nu-n)}{2N}$
$\tilde{S}_2'$	$\frac{\sum(\mu+m+\nu+n-2T(\mu,n)-2T(m,\nu))}{N-\sum(T(\mu,n)+T(m,\nu))+max\{\sum(\mu+n),\sum(m+\nu)\}}$

Table A2. *Cont.*

Similarity Measure	Expression
$\tilde{S}_6'$	$\frac{\sum(\mu+m+\nu+n-2T(\mu,n)-2T(m,\nu))}{2N}$
$\tilde{S}_1^c$	$\frac{N+\sum(T(\mu,m)+T(\nu,n)-\mu-m}{N-min\{\sum(\mu,\nu),\sum(m,n)\}}$
$\tilde{S}_5^c$	$\frac{N+\sum(T(\mu,m)+T(\nu,n)-\nu-n}{N-\sum(\nu+n-T(\mu,m)-T(\nu,n))}$

Table A3. Mathematical expressions of the reviewed similarity measures III.

Similarity Measure	Expression
L_5	$1-\sqrt[\rho]{\frac{\sum\left(\frac{\Delta_\mu+\Delta_\nu}{\mu\vee m+\nu\vee n}\right)^\rho}{N}}$
N_6	$1-\sqrt[\rho]{\frac{\sum\left(\frac{\Delta_\mu+\Delta_\nu}{2}\right)^\rho}{N}}$
F_5	$\frac{e^{2-\sqrt[\rho]{\frac{\sum(\Delta_\mu)^\rho}{N}}\sqrt[\rho]{\frac{\sum(\Delta_\nu)^\rho}{N}}}-1}{e^2-1}$
F_6	$1-\frac{\left(\sqrt[\rho]{\frac{\sum(\Delta_\mu)^\rho}{N}}\right)^u+\left(\sqrt[\rho]{\frac{\sum(\Delta_\nu)^\rho}{N}}\right)^v}{2},(u,v>0)$
F_7	$1-\frac{\sin\frac{\pi_{3.14}}{2}\left(\sqrt[\rho]{\frac{\sum(\Delta_\mu)^\rho}{N}}\right)+\left(\sin\frac{\pi_{3.14}}{2}\sqrt[\rho]{\frac{\sum(\Delta_\nu)^\rho}{N}}\right)}{2}$
S_{WY}	$\frac{1}{2}\sum w_i\left(\sqrt{\mu*m}+2\sqrt{\nu*n}+\sqrt{\pi*p}+\sqrt{(1-\nu)(1-n)}\right)$
S_{CCL}	$\sum\left(w_i\left(1-\frac{\lvert 2(\mu-m)-(\nu-n)\rvert}{3}\left(1-\frac{\pi+p}{2}\right)-\frac{\lvert 2(\nu-n)-(\mu-m)\rvert}{3}\left(\frac{\pi+p}{2}\right)\right)\right)$
W_{IFSSs}	$\frac{\sum w_i((\mu*m)+(\nu*n))}{\sum((\mu^2\vee m^2)+(\nu\vee n)^2)}\left(\sum w_i\right)$

Table A4. Mathematical expressions of the reviewed distance measures.

Distance Measure	Expression
d_e	$\sqrt{\frac{1}{2}\sum_{i=1}^{n}\left[\Delta_\mu^2+\Delta_\nu^2\right]}$
d_h	$\frac{1}{2}\sum_{i=1}^{n}[\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert]$
d_{ne}	$\sqrt{\frac{1}{2n}\sum_{i=1}^{n}\left[\Delta_\mu^2+\Delta_\nu^2\right]}$
d_{nh}	$\frac{1}{2n}\sum_{i=1}^{n}[\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert]$
d_{IFS}^1	$\sqrt{\frac{1}{2}\sum_{i=1}^{n}\left[\Delta_\mu^2+\Delta_\nu^2+\Delta_\pi^2\right]}$
e_{IFS}^1	$\frac{1}{2}\sum_{i=1}^{n}[\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert+\lvert\Delta_\pi\rvert]$
l_{IFS}^1	$\sqrt{\frac{1}{2n}\sum_{i=1}^{n}\left[\Delta_\mu^2+\Delta_\nu^2+\Delta_\pi^2\right]}$

Table A4. *Cont.*

Distance Measure	Expression
q_{IFS}^{1}	$\frac{1}{2n}\sum_{i=1}^{n}[\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert+\lvert\Delta_\pi\rvert]$
d_{h2}	$\sum \max\{\lvert\Delta_\mu\rvert,\lvert\Delta_\nu\rvert\}$
e_h	$\frac{1}{N}\sum \max\{\lvert\Delta_\mu\rvert,\lvert\Delta_\nu\rvert\}$
l_h	$\sqrt{\sum \max\{\Delta_\mu^2,\Delta_\nu^2\}}$
q_h	$\sqrt{\frac{1}{N}\sum \max\{\Delta_\mu^2,\Delta_\nu^2\}}$
d_1	$\sum_{i=1}^{n} w_i\left[\frac{\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert}{4}+\frac{\max\{\lvert\Delta_\mu\rvert,\lvert\Delta_\nu\rvert\}}{2}\right]/\sum_{i=1}^{n} w_i$
d_2^p	$\frac{1}{\sqrt[\rho]{n}}\sqrt[\rho]{\sum_{i=1}^{n}\left(\frac{\lvert\Delta_\mu\rvert+\lvert\Delta_\nu\rvert}{2}\right)^\rho}$
d_{eh}	$\sum_{i=1}^{n}\max\{\lvert\Delta_\mu\rvert,\lvert\Delta_\nu\rvert,\lvert\Delta_\pi\rvert\}$
l_{eh}	$\frac{1}{N}\sum_{i=1}^{n}\max\{\lvert\Delta_\mu\rvert,\lvert\Delta_\nu\rvert,\lvert\Delta_\pi\rvert\}$
l_{eh}	$\sqrt{\sum \max\{\Delta_\mu^2,\Delta_\nu^2,\Delta_\pi^2\}}$
q_{eh}	$\sqrt{\frac{1}{N}\sum \max\{\Delta_\mu^2,\Delta_\nu^2,\Delta_\pi^2\}}$
D_{IFS}	$I_{IFS}(A,B)+I_{IFS}(B,A)$, where $I_{IFS}(A,B)=\sum\left[\mu\ln\frac{\mu}{\frac{1}{2}(\mu+m)}+\nu\ln\frac{\nu}{\frac{1}{2}(\nu+n)}\right]$

References

1. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
2. Liu, B. Sentiment Analysis and Opinion Mining. 2012. Available online: https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf (accessed on 14 March 2022).
3. Řehůřek, R.; Kolkus, M. *Language Identification on the Web: Extending the Dictionary Method*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 357–368. [CrossRef]
4. Cai, W.; Cai, D.; Huang, S.; Li, M. Utterance-Level End-to-End Language Identification Using Attention-Based CNN-BLSTM. 2019. Available online: https://arxiv.org/abs/1902.07374 (accessed on 14 March 2022).
5. Brenes, D.J.; Gayo-Avello, D.; Pérez-González, K. Survey and Evaluation of Query Intent Detection Methods. In Proceedings of the 2009 Workshop on Web Search Click Data WSCD '09, Barcelona, Spain, 9–11 February 2009; Association for Computing Machinery: New York, NY, USA, 2009; pp. 1–7. [CrossRef]
6. Wang, C.; Blei, D.M. *Collaborative Topic Modeling for Recommending Scientific Articles*; ACM Press: New York, NY, USA, 2011; p. 448. [CrossRef]
7. Zadeh, L. Fuzzy sets. *Inf. Control.* **1965**, *8*, 338–353. [CrossRef]
8. Małysiak-Mrozek, B. Uncertainty, imprecision, and many-valued logics in protein bioinformatics. *Math. Biosci.* **2019**, *309*, 143–162. [CrossRef] [PubMed]
9. Atanassov, K.T. Intuitionistic Fuzzy Sets. *Fuzzy Sets Syst.* **1986**, *20*, 87–96. [CrossRef]
10. Intarapaiboon, P. *A Framework for Text Classification Using Intuitionistic Fuzzy Sets*; Springer: Berlin/Heidelberg, Germany, 2019 ; pp. 737–746. [CrossRef]
11. Abbasi, B.Z.; Hussain, S.; Faisal, M.I. *An Automated Text Classification Method: Using Improved Fuzzy Set Approach for Feature Selection*; IEEE: Islamabad, Pakistan, 2019; p. 5.
12. Widyantoro, D.H.; Yen, J. A Fuzzy Similarity Approach in Text Classification Task. 1998. Available online: https://www.researchgate.net/publication/3845652_A_fuzzy_similarity_approach_in_text_classification_task (accessed on 14 March 2022).
13. Puri, S. A Fuzzy Similarity Based Concept Mining Model for Text Classification. 2012. Available online: https://arxiv.org/ftp/arxiv/papers/1204/1204.2061.pdf (accessed on 14 March 2022).
14. Jiang, J.Y.; Liou, R.J.; Lee, S.J. A fuzzy self-constructing feature clustering algorithm for text classification. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 335–349. [CrossRef]
15. Jiang, J.Y.; Tsai, S.C.; Lee, S.J. Expert Systems with Applications FSKNN: Multi-label text categorization based on fuzzy similarity and k nearest neighbors q. *Expert Syst. Appl.* **2012**, *39*, 2813–2821. [CrossRef]

16. Szmidt, E.; Kacprzyk, J. Using Intuitionistic Fuzzy Sets in Text Categorization. In *Artificial Intelligence and Soft Computing–ICAISC 2008*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 351–362. [CrossRef]
17. Bounabi, M.; Moutaouakil, K.E.; Satori, K. *Text Classification Using Fuzzy TF-IDF and Machine Learning Models*; ACM: Rabar, Morocco, 2019; p. 6.
18. Behera, B. Text Document Classification Using Fuzzy Rough Set Based on Robust Nearest Neighbor (FRS-RNN). *Soft Comput.* **2021**, *25*, 9915–9923. [CrossRef]
19. Puri, S. Efficient Fuzzy Similarity-Based Text Classification with SVM and Feature Reduction. In Proceedings of the Congress on Intelligent Systems, New Delhi, India, 5–6 September 2020; Sharma, H., Saraswat, M., Yadav, A., Kim, J.H., Bansal, J.C., Eds.; Springer: Singapore, 2020; pp. 341–356.
20. Wang, W.; Xin, X. Distance measure between Intuitionistic Fuzzy Sets. *Pattern Recognit. Lett.* **2005**, *26*, 2063–2069. [CrossRef]
21. Kacprzyk, J. *Multistage Fuzzy Control: A Prescriptive Approach*, 1st ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1997.
22. Atanassov, K.T. Intuitionistic Fuzzy Sets. In *Studies in Fuzziness and Soft Computing*; Physica-Verlag HD Fuzzy Sets and Systems: Berlin/Heidelberg, Germany, 1999; Volume 35. [CrossRef]
23. Szmidt, E.; Kacprzyk, J. Distances between Intuitionistic Fuzzy Sets. *Fuzzy Sets Syst.* **2000**, *114*, 505–518. [CrossRef]
24. Li, D.; Cheng, C. New Similarity Measures of Intuitionistic Fuzzy Sets and Application to Pattern Recognitions. *Pattern Recognit. Lett.* **2002**, *23*, 221–225. [CrossRef]
25. Mitchell, H. On the Dengfeng–Chuntian Similarity Measure and Its Application to Pattern Recognition. *Pattern Recognit. Lett.* **2003**, *24*, 3101–3104. [CrossRef]
26. Hung, W.L.; Yang, M.S. Similarity Measures of Intuitionistic Fuzzy Sets Based on Hausdorff Distance. *Pattern Recognit. Lett.* **2004**, *25*, 1603–1611. [CrossRef]
27. Chen, S.M. Measures of similarity between vague sets. *Fuzzy Sets Syst.* **1995**, *74*, 217–223. [CrossRef]
28. Sidiropoulos, G.K.; Apostolidis, K.D.; Damianos, N.; Papakostas, G.A. Fsmpy: A Fuzzy Set Measures Python Library. *Information* **2022**, *13*, 64. [CrossRef]
29. Grzegorzewski, P. Distances between Intuitionistic Fuzzy Sets and/or Interval-Valued Fuzzy Sets Based on the Hausdorff Metric. *Fuzzy Sets Syst.* **2004**, *148*, 319–328. [CrossRef]
30. Chen, S.-M. Similarity Measures between Vague Sets and between Elements. *IEEE Trans. Syst. Man Cybern. Part Cybern.* **1997**, *27*, 153–158. [CrossRef]
31. Hong, D.H.; Kim, C. A Note on Similarity Measures between Vague Sets and between Elements. *Inf. Sci.* **1999**, *115*, 83–96. [CrossRef]
32. Liang, Z.; Shi, P. Similarity Measures on Intuitionistic Fuzzy Sets. *Pattern Recognit. Lett.* **2003**, *24*, 2687–2693. [CrossRef]
33. Liu, H.W. New Similarity Measures between Intuitionistic Fuzzy Sets and between Elements. *Math. Comput. Model.* **2005**, *42*, 61–70. [CrossRef]
34. Zhang, C.; Fu, H. Similarity Measures on Three Kinds of Fuzzy Sets. *Pattern Recognit. Lett.* **2006**, *27*, 1307–1317. [CrossRef]
35. Hung, W.L.; Yang, M.S. Similarity Measures of Intuitionistic Fuzzy Sets Based on Lp Metric. *Int. J. Approx. Reason.* **2007**, *46*, 120–136. [CrossRef]
36. Park, J.H.; Park, J.S.; Kwun, Y.C.; Lim, K.M. New Similarity Measures on Intuitionistic Fuzzy Sets. In *Fuzzy Information and Engineering*; Cao, B.Y., Ed.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 40, pp. 22–30. [CrossRef]
37. Hung, W.L.; Yang, M.S. On Similarity Measures between Intuitionistic Fuzzy Sets. *Int. J. Intell. Syst.* **2008**, *23*, 364–383. [CrossRef]
38. Hung, W.L.; Yang, M.S. On the J-divergence of Intuitionistic Fuzzy Sets with Its Application to Pattern Recognition. *Inf. Sci.* **2008**, *178*, 1641–1650. [CrossRef]
39. Ye, J. Cosine Similarity Measures for Intuitionistic Fuzzy Sets and Their Applications. *Math. Comput. Model.* **2011**, *53*, 91–97. [CrossRef]
40. Hwang, C.M.; Yang, M.S. Modified Cosine Similarity Measure between Intuitionistic Fuzzy Sets. In *Artificial Intelligence and Computational Intelligence*; Lei, J., Wang, F., Deng, H., Miao, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7530, pp. 285–293.
41. Julian, P.; Hung, K.C.; Lin, S.J. On the Mitchell Similarity Measure and Its Application to Pattern Recognition. *Pattern Recognit. Lett.* **2012**, *33*, 1219–1223. [CrossRef]
42. Iancu, I. Intuitionistic Fuzzy Similarity Measures Based on Frank T-Norms Family. *Pattern Recognit. Lett.* **2014**, *42*, 128–136. [CrossRef]
43. Deng, G.; Jiang, Y.; Fu, J. Monotonic Similarity Measures between Intuitionistic Fuzzy Sets and Their Relationship with Entropy and Inclusion Measure. *Inf. Sci.* **2015**, *316*, 348–369. [CrossRef]
44. Song, Y.; Wang, X.; Lei, L.; Xue, A. A Novel Similarity Measure on Intuitionistic Fuzzy Sets with Its Applications. *Appl. Intell.* **2015**, *42*, 252–261. [CrossRef]
45. Chen, S.M.; Cheng, S.H.; Lan, T.C. A Novel Similarity Measure between Intuitionistic Fuzzy Sets Based on the Centroid Points of Transformed Fuzzy Numbers with Applications to Pattern Recognition. *Inf. Sci.* **2016**, *343–344*, 15–40. [CrossRef]
46. Muthukumar, P.; Sai Sundara Krishnan, G. A Similarity Measure of Intuitionistic Fuzzy Soft Sets and Its Application in Medical Diagnosis. *Appl. Soft Comput.* **2016**, *41*, 148–156. [CrossRef]
47. Yang, Y.; Chiclana, F. Consistency of 2D and 3D Distances of Intuitionistic Fuzzy Sets. *Expert Syst. Appl.* **2012**, *39*, 8665–8670. [CrossRef]

48. Vlachos, I.K.; Sergiadis, G.D. Intuitionistic Fuzzy Information–Applications to Pattern Recognition. *Pattern Recognit. Lett.* **2007**, *28*, 197–206. [CrossRef]
49. Nguyen, H. A Novel Similarity/Dissimilarity Measure for Intuitionistic Fuzzy Sets and Its Application in Pattern Recognition. *Expert Syst. Appl.* **2016**, *45*, 97–107. [CrossRef]
50. Greene, D.; Cunningham, P. Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering. In Proceedings of the 23rd International Conference on Machine Learning (ICML'06), Pittsburgh, PA, USA, 25–29 June 2006; pp. 377–384.
51. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G. Scikit-Learn: Machine Learning in Python. Available online: https://arxiv.org/abs/1201.0490 (accessed on 11 August 2021).
52. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python*, 1st ed.; O'Reilly: Newton, MA, USA, 2009.
53. Hatzimichailidis, A.; Papakostas, G.; Kaburlasos, V. A novel distance measure of Intuitionistic Fuzzy Sets and its application to pattern recognition problems. *Int. J. Intell. Syst.* **2012**, *27*, 396–409. [CrossRef]
54. Xu, Z. Some Similarity Measures of Intuitionistic Fuzzy Sets and Their Applications to Multiple Attribute Decision Making. *Fuzzy Optim. Decis. Mak.* **2007**, *6*, 109–121. [CrossRef]