Data & Knowledge Engineering 158 (2025) 102421

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

A graph theoretic approach to assess quality of data for classification task

Payel Sadhukhan [a], Samrat Gupta [b,c,d,*]

[a] Department of Computer Science and Engineering (IoT) Techno Main, Salt Lake, WB, India
[b] Information Systems Area, Indian Institute of Management, Ahmedabad, India
[c] Department of Finance, University of Economics in Bratislava, Slovakia
[d] Department of Information Systems, University of Agder, Kristiansand, Norway

ARTICLE INFO

Keywords:
Data quality
Classification
Graph theory
Data marketplace
Feature importance

ABSTRACT

The correctness of predictions rendered by an AI/ML model is key to its acceptability. To foster researchers' and practitioners' confidence in the model, it is necessary to render an intuitive understanding of the workings of a model. In this work, we attempt to explain a model's working by providing some insights into the quality of data. While doing this, it is essential to consider that revealing the training data to the users is not feasible for logistical and security reasons. However, sharing some interpretable parameters of the training data and correlating them with the model's performance can be helpful in this regard. To this end, we propose a new measure based on Euclidean Minimum Spanning Tree (EMST) for quantifying the intrinsic separation (or overlaps) between the data classes. For experiments, we use datasets from diverse domains such as finance, medical, and marketing. We use state-of-the-art measure known as *Davies Bouldin Index (DBI)* to validate our approach on four different datasets from aforementioned domains. The experimental results of this study establish the viability of the proposed approach in explaining the working and efficiency of a classifier. Firstly, the proposed measure of class-overlap quantification has shown a better correlation with the classification performance as compared to DBI scores. Secondly, the results on multi-class datasets demonstrate that the proposed measure can be used to determine the feature importance so as to learn a better classification model.

1. Introduction

The ever-lasting human urge to render their decision-making capabilities to a machine has resulted in significant advances in artificial intelligence and machine learning [1,2]. *Automated decision making* – a natural extension of this development has emerged as an important area of research and practice. The hardware and software competencies associated with the operation of a model are improving with each passing day. This has enabled the models to handle data dimensions beyond normal human perception. However, this phenomenon often creates an opacity in the model's operation for the users receiving the service [3]. Additionally, it has been observed that in some cases, semantically incorrect or irrelevant reasoning drove the decisions of a computationally efficient model [4]. It is found that the useful revelations from a given data are often camouflaged with the irrelevant ones [5]. This calls for an increasing need for an understanding of the quality of data and the working of the automated decision-making models [6,7].

[*] Corresponding author at: Information Systems Area, Indian Institute of Management, Ahmedabad, India.
E-mail address: samratg@iima.ac.in (S. Gupta).

https://doi.org/10.1016/j.datak.2025.102421
Received 11 June 2024; Received in revised form 30 December 2024; Accepted 15 February 2025

Available online 3 March 2025

An automated decision-making model is built or crafted on the training data (*training phase*) wherein the model's task is to predict an unseen test data from the learning of the *training phase*. We may note that we can *understand, explain, or interpret* a model through the data on which it is trained. Following this line of thought, we argue that we can evaluate the efficiency of a model by interpreting the intrinsic class separations (or class overlaps) of the data on which it is trained.

To this end, we pose the following research questions:

RQ1: Can we understand the outcomes of a model by assessing the quality of the data on which it is trained?

In this research, we work on intuiting the modus operandi of a model by exploring the training data. The key element of our exploration is the recognition and quantification of the overlaps of the categories (or classes) in the data and correlating them with the predictive power of the model. There are usually two or more classes in a dataset, and we work on perceiving (quantitatively) the separations (or overlaps) between the classes that exist in the dataset. An efficient model can be built from data that has well-separated classes as compared to the one with overlapping classes [8]. For example, we can succeed in building a model to distinguish between *fraudulent* and *authentic* bank transactions, if the sets of fraudulent and authentic transactions differ from each other in several characteristics (well separated from each other). In this work, we determine the overlap (or separation) in the training data so as to help the developers build better models and users in explaining the working and performance of a model. In doing so, we propose a Euclidean Minimum Spanning Tree (EMST) based novel measure to quantify the class-overlap intrinsic to a dataset. We have formulated the design of this measure on the basis of two key aspects. First, we convert the set of points in a dataset into a connected graph by constructing their EMST. Subsequently, the proportion of homogeneous edges and heterogeneous edges in the transformed graph, and the weights of the homogeneous and the heterogeneous edges. We also use the Davies Bouldin Index (DBI), a popular and existing measure used in the field of data clustering for the purpose of benchmarking [9].

A model trained on the dataset having overlapping classes (has regions where points from two or more distinct classes co-exist), will have difficulty in classifying points which lie in the overlapped regions. In this work, the *EMST Overlap Index* (EMSTOI) and DBI of a set of datapoints are used to estimate the separations of the intrinsic classes present in it. Usually, a dataset with well-separated classes will possess lower EMSTOI and lower DBI than the one with overlapping classes. This information on intrinsic class separations in the training data can be shared with the practitioners, developers, and users of the model (trained on this data). A lower EMSTOI or DBI (indicating lesser class overlap) can foster the users' confidence in the outcomes and predictions of the model. It can also help them in distinguishing instrumental features from the less expressive ones. Further, it can help the users, practitioners, and developers in handling critical data types like imbalanced data [10,11]. Sharing this information is a better option (as compared to sharing the raw and original data with the users) in terms of logistics as well as privacy and integrity.

We conduct experimental studies to explore class overlap in datasets originating from diverse domains:

In the first study, we used the overlap information obtained from EMSTOI and DBI to ascertain the quality of balanced datasets. We conduct experimental studies on datasets from finance domain. The first dataset in this category known as PaySim is a dataset related to credit card transactions, while the second dataset referred as BankSim is related to the authenticity of bank transactions. These are heavily class-imbalanced synthetic datasets consisting of fraudulent and authentic transactions. We undersample the majority class (set of authentic transactions) to get balanced datasets. We repeat this process several times. We estimate the class separations in the balanced sets and show that in both cases, classifiers render superior performance when the EMSTOI or DBI in the balanced set is low, wherein the proposed EMSTOI is superior to DBI. This establishes the utility of the proposed EMSTOI in assessing the quality of data and subsequently in explaining the model. As a part of this study, we conducted two more experiments to study the relationship between the average amount saved by detecting the fraudulent transactions and class overlap indices (EMSTOI and DBI) and the average amount lost by missing the detection of fraudulent transactions and class overlap indices (EMSTOI and DBI). We observed a negative correlation between the average amount saved by fraudulent transaction detection with EMSTOI and DBI. More amount was saved when the classifier was trained on data with well-separated classes (low EMSTOI and DBI). It was further seen that more amount was lost (by missing the detection of fraudulent transactions) when the model was trained on data with overlapping classes (high EMSTOI and DBI).

In the second study, we determine feature importance using the EMSTOI and DBI, followed by the incorporation of computed weights into the classifier modeling, wherein this augmentation improved the learning of the classifier. First, we consider a multi-class dataset from medical domain which pertains to maternal health. In this dataset each data point contains information on the age and five clinical parameters of pregnant women. Second, we consider multi-class dataset originating from marketing domain. This dataset reports the advertising expenditure on different media such as television, radio, and newspaper. We explore the importance of each of the features of these datasets through their EMSTOI and DBI. We trained several classifiers, each modeled and dedicated to a single feature. The performance of the dedicated classifiers is correlated with the EMSTOI and DBI of the individual features. The lower the EMSTOI and DBI for a feature, the better the classification accuracy. This motivated us to conduct another set of experiments where we build an enhanced model incorporating the feature importance. The classification performance of the enhanced model was superior to that of the base model.

The two aforementioned studies demonstrate that the developers, practitioners, and users of a model can use the EMSTOI and DBI of training data to understand and explain the workings of a model. These scores are one-dimensional information which can be shared without logistic or privacy concerns. Another advantage of the proposed approach is its model agnostic understanding based on the training data. The proposed EMSTOI can find utility in data marketplaces where the similarity among datasets needs to be derived for enabling functionalities such as data valuation and revenue allocation [12,13]. EMSTOI can contribute to quantifying the intrinsic characteristics of datasets to derive their value for machine learning tasks. A data set that has low overlap (in terms of EMSTOI) has a potential to train a better model and can be more valuable for a buyer in the prediction of business outcomes. Class overlap which is an intrinsic characteristic of a dataset (and can be measured through EMSTOI) can be combined with its extrinsic characteristics such as environmental sustainability and perceived uniqueness to recommend its price to the seller while publishing the dataset on a data marketplace.

The contributions of this work are as follows:

- We propose an approach to understand the working of a model by exploring data quality. This investigation is performed in the pre-modeling phase and the learning does not depend on the type of classifier model. Such an approach offers an advantage over other feature-explainability approaches [14–16] which generate the explanation after the prediction phase only.
- We design a novel measure based on the Euclidean Minimum Spanning Tree (EMST) to recognize and quantify the degree of class overlap in a dataset. We benchmark the proposed measure with DBI, a popular existing measure in the domain of data clustering. The proposed measure named as EMSTOI outperforms state-of-the-art DBI.
- The proposed approach can be utilized to learn the feature importance as well as the overall quality of the training data. The working efficiency of a model can be explained through this information.
- Experimental studies demonstrate that the proposed EMSTOI can uncover many industry-oriented aspects, such as the amount of money saved by detecting fraudulent transactions, the amount of money lost by missing the detection of fraudulent transactions, and developing an enhanced classifier by incorporating the learned feature importance.

The rest of this article is organized as follows. In the next section, we discuss the extant work on model explainability and data quality. The proposed approach is presented in the following section. The subsequent section outlines the experimental setup of this study. The results of the experimental analysis are presented next. Following this, we discuss the theoretical and managerial implications as well as the limitations of this study. Finally, we conclude this work.

2. Background and related work

2.1. Explainability of data-driven models

In the past decade, AI-based data-driven models have been increasingly used in industries such as finance and banking [17,18], recruitment and selection in education [19], resource utilization in healthcare [20]. These models have also been applied to prevent the spread of misinformation [21], improve the resilience of supply chains [22,23], and enhance business outcomes of movies [24]. This remarkable upsurge of AI has given rise to a follow-up question of how to explain the decisions provided by the automated systems [25].

To understand the causation leading to the emergence of explainability of data-driven models, we need to briefly recapitulate the timeline of AI developments over the past few decades. The initial years of development in the field of AI (1950–1960) were characterized by rule-based learning systems which were easy to interpret but had low accuracy and predictive power thus limiting their utility [26]. In 1980s researchers began focusing on improving the predictive powers of AI models [27]. This led to the emergence of statistical learning methods popularly known as Machine Learning [28]. Subsequently, in 1990s and early 2000s, betterment in accuracy was further achieved by more complex models which were based on artificial neural networks and ensemble methods [29,30]. However, this progression came with a cost wherein as the performance quality of the models increased, the working of the models became increasingly murky and non-explainable to the users [31]. For instance, a neural-network based model for distinguishing husky images and wolves images delivered commendable accuracy on the task, but the model only captured and provided outcome on the basis of background snow instead of the characteristic difference of the two animals [32]. Additionally, the quantitative role of prior information in data leads to some bias in the decision-making rationale of a system [33] . Hence, data-driven explainability comes with substantial new challenges and opportunities [34,35] .

One approach for providing of a data-driven model operates on a post-hoc basis. This approach evaluates the explainability of a model on the basis of the predictions from the model. Post-hoc explainability approaches are also used to decipher the modus operandi of an opaque model by the use of rules [36,37], anchors [38] and surrogates [39,40]. Partial Dependence Plots (PDP) [41], Accumulated Local Effect (ALE) Plots [42] and Individual Conditional Expectations (ICE) [43] constitute this class of approach. There is another approach for explaining the predictions wherein specific methods explain the role of the features to arrive at a decision. Some examples of such specific methods are LIME [14], and SHAP [16]. There are several other approaches to explain the features such as, analyzing the performance after permutation of features [44], and counterfactual explanations of the decisions [45,46].

2.2. Assessing data quality

The output rendered by a model not only depends on the input which is being fed but also on the data on which it has been trained. Following this line of reasoning, we focus on understanding the data which trains the model providing an ante-hoc explanation of the decisions to be taken by a model. Usually, some intrinsic characteristics of a dataset such as, dispersion, separability, and class imbalance can affect the goodness and working of a model. From a technical perspective, diverse models such as Decision Trees, Support Vector Machines, Neural Network and Nearest Neighbor-based Classifiers have different modus operandi and can operate differently on the same training dataset. However, a part of this training could be attributed to the quality of the training data. Consequently, knowledge of the dataset can provide us considerable insights into the working of different models. Hence, in addition to the explainability of data-driven models, the practitioners are often posed with the challenge of choosing the right data for a task. A few works in literature address similar and related concerns. For example, [47] looks for interesting aspects to be explored in a data mining task, [48] focuses on finding the right metric to bring out information and [49] resorts to domain-guided questionnaire to obtain the right kind of data.

Now, we discuss the utility of a popular index which quantifies the separation between the collections of points present in a feature space. In machine learning, clustering is a popular methodology for grouping the data points where similar points are added into the same group while dissimilar points belong to different groups [50]. Several popular measures exist, which estimate the goodness of the clustering task [51–53]. One such measure is DBI [9]. Though it has been a popular choice and extensively used for evaluating the goodness of clustering output, we didn't find an extant work where it has been used to evaluate the separation of a set of data-points belonging to different classes. This approach of evaluation can be used for a given set of data-points and we use this line of action in our work. The lesser the value of DBI obtained in a task, the more is the separation among the classes.

In this work, we demonstrate the use of DBI to estimate the intrinsic separations existing between the classes present in the training data. While computing DBI for the training data, we assume that it is a collection of c distinct classes. The computation of DBI is dependent on the similarities between the points arising from different classes. For each class i, $1 \leq i \leq c$, the computation of DBI involves finding out quantitative value of its similarity with another class j, $1 \leq j \leq c$, $j \neq i$. DBI is the cumulative sum of such quantitative similarities overall i, that is for all the classes. The mathematical definition and formulation of DBI is provided in Appendix A.1. For a model which is assigned with the task of classifying the data-points, it is desirable that the classes in the training data should be distinct from each other. In mathematical terminology, the different categories should originate from distributions with highly different population means [54] thus resulting in a low DBI. The minimum value of DBI is *zero* and denotes the ideal scenario of maximum separation between the categories. However, DBI provides overlap quantification under several stringent assumptions such as the classes should possess disparate cluster centers. As such, DBI cannot be relied upon to perceive the proportion of class overlap in a dataset. Hence, we need a more generalized index to quantify class overlaps while providing better explainability into the goodness of a dataset. Leveraging this gap in the research we propose a new measure based on the Euclidean Minimum Spanning Tree for quantifying the class overlaps in the training data. Since data quality particularly in classification tasks is a function of intrinsic characteristics of a dataset it is important to investigate class overlap in a dataset. The objective of this study is to investigate how the working of a model can be explained without generating the model itself such that the explanation is model-agnostic. This is based on the intuition that training data that has good intrinsic separation of its classes will result in the implementation of an efficient classifier. On the contrary, training data with overlapping classes will generate a sub-optimal classifier [55]. In order to validate our premise, we correlate quantified class overlap indices with the predictions of different models. In doing so, we also benchmark the proposed EMSTOI with state-of-the-art DBI to determine the intrinsic overlaps in the training data.

3. Proposed research methodology

An insight into the overlap of data-points in the training data can explain the impending classification performance by determining the goodness of the undersampled dataset as well as feature importance. We work towards this objective and propose a novel measure to quantify the class overlaps in a dataset. As mentioned in section 2.2, the new measure overcomes the limitations of state-of-the-art DBI. DBI can provide information about class overlap only when the classes originate from distinct clusters. It can provide a sub-optimal output when the origin of the classes in a dataset overlaps. We quantify and compute the separations intrinsic to a dataset by using the proposed EMSTOI and DBI in the pre-modeling phase. In the post-modeling phase, we obtain the classes of a set of unknown test points from the model and compute the classification performance (accuracy and F_1). Subsequently, we compute the correlation between the class overlap scores and classification performance with an expectation that higher separations between the classes present in a dataset (lower class overlap index) will be positively correlated with classification performance.

3.1. A new measure based on minimum spanning tree

We use Euclidean Minimum Spanning Tree (EMST) based approach to form a connected network from a given set of datapoints. An EMST of a finite set of n points in a feature space connects them by a set of $n - 1$ line segments where the points serve as the endpoints, minimizing the total length of the segments. The technical foundation of EMST is provided in Appendix A.2. It is important to note that, each of the n points can belong to any one of the given classes. After forming the EMST, we look at each edge. If both the vertices of an edge belong to the same class, we call it a *homogeneous edge*. If the vertices of an edge belong to different classes, we call it a *heterogeneous edge*. The percentage of heterogeneous edges in an EMST indicate the degree of class overlap in a given dataset. Further, the average weights of the homogeneous edges and the heterogeneous edges also provide insight into the class overlaps. In a dataset with low-class overlap, each class will be tight-knit resulting in low average homogeneous edge weights and a higher value for average heterogeneous edge weight. We employ a ratio of the number of heterogeneous edges over the number of homogeneous edges, and ratio of the average weight of homogeneous edges over the average weight of heterogeneous edges to quantify the overlap of the classes present in a dataset. We define the EMST-based class-overlap of a given dataset $\mathcal{D}$ denoted by EMSTO as follows:

$$\text{EMSTO} = \frac{\varepsilon_{hom}}{\varepsilon_{het}} \times \frac{\nu_{het}}{\nu_{hom}} \tag{1}$$

where ε_{hom} and ε_{het} denote the average homogeneous edge weight and average heterogeneous edge weight respectively. ν_{hom} and ν_{het} denote the number of homogeneous edges and the number of heterogeneous edges in the EMST of the given dataset. The technical details of the proposed equation are detailed in Appendix A.3. The more the number of heterogeneous edges, the more is overlap between the classes. Additionally, smaller homogeneous edge weights (shorter homogeneous edges) indicate compact class structures. A low-class overlap wherein homogeneous edges have lower weights than the heterogeneous edges is desirable as the classes therein are separated from each other and it is expected that the models trained on such a dataset would deliver efficient performance. Thus, we compute and explain the class separations intrinsic to a dataset and, consequently, to the model. The values of EMSTO can range from 0 to ∞. The former indicates no overlap while the latter is obtained when the classes originate from identical distribution. When we want to compare two distributions, such diverse range of values can be too abstract. To address this shortcoming, we tweak the EMSTO value mathematically in the following manner to compute EMST based class overlap index (EMSTOI). Similar to DBI, this index quantifies the overlap in a dataset in a more effective way.

$$\text{EMSTOI} = \frac{\text{EMSTO}}{\text{EMSTO} + 1} \tag{2}$$

Unlike EMSTO, the range of EMSTOI is bounded in $[0, 1]$. The higher the overlap, the higher the value of EMSTOI. The mathematical properties of EMSTOI are discussed in Appendix A.4. Fig. 1 shows the variation of EMSTOI and DBI in differently overlapped datasets. In each sub-figure of Fig. 1, we have constructed a synthetic 2-class dataset where the points belong to exactly one of the two given classes (indicated by blue and orange colors). For a scenario wherein each class originates from a single cluster, both EMSTOI and DBI have shown strong correlations with the degree of overlap. The values of both measures have increased with the increase in the class overlap.

However, in situations where each class originates from multiple clusters, DBI can fail to correctly capture the overlap in the data. The issue happens when the cluster centers of different classes are located within a small vicinity. The technical details can be found in Appendix A.1. Fig. 2 shows four dataset in which each class originates from two or more non-overlapping clusters. Consequently, the classes in each dataset are well-separated. In all the cases, EMSTOI scores are approximately 0, and it is in congruence with the actual scenario as there is a low overlap between the classes. On the contrary, the value of DBI is quite high, spuriously indicating a strong overlap between the classes. This anomaly in the DBI arises because the resultant means (cluster centers) for the two classes are collocated thus misguiding the DBI computation. The proposed measure EMSTOI is immune to such variations and can render correct indications about the class overlap present in the data.

3.2. Addressing the class imbalance problem

Class-imbalance problem is a conspicuous characteristic of data arising from a number of real-world domains [56,57]. The learning of models gets severely plagued due to this issue [58]. Researchers and practitioners often work on balancing the classes of imbalanced datasets. The most popular and convenient way to balance a dataset is *undersampling* which means eliminating the points from the majority class. However, undersampling can be effective only if it provides a separation of the classes.

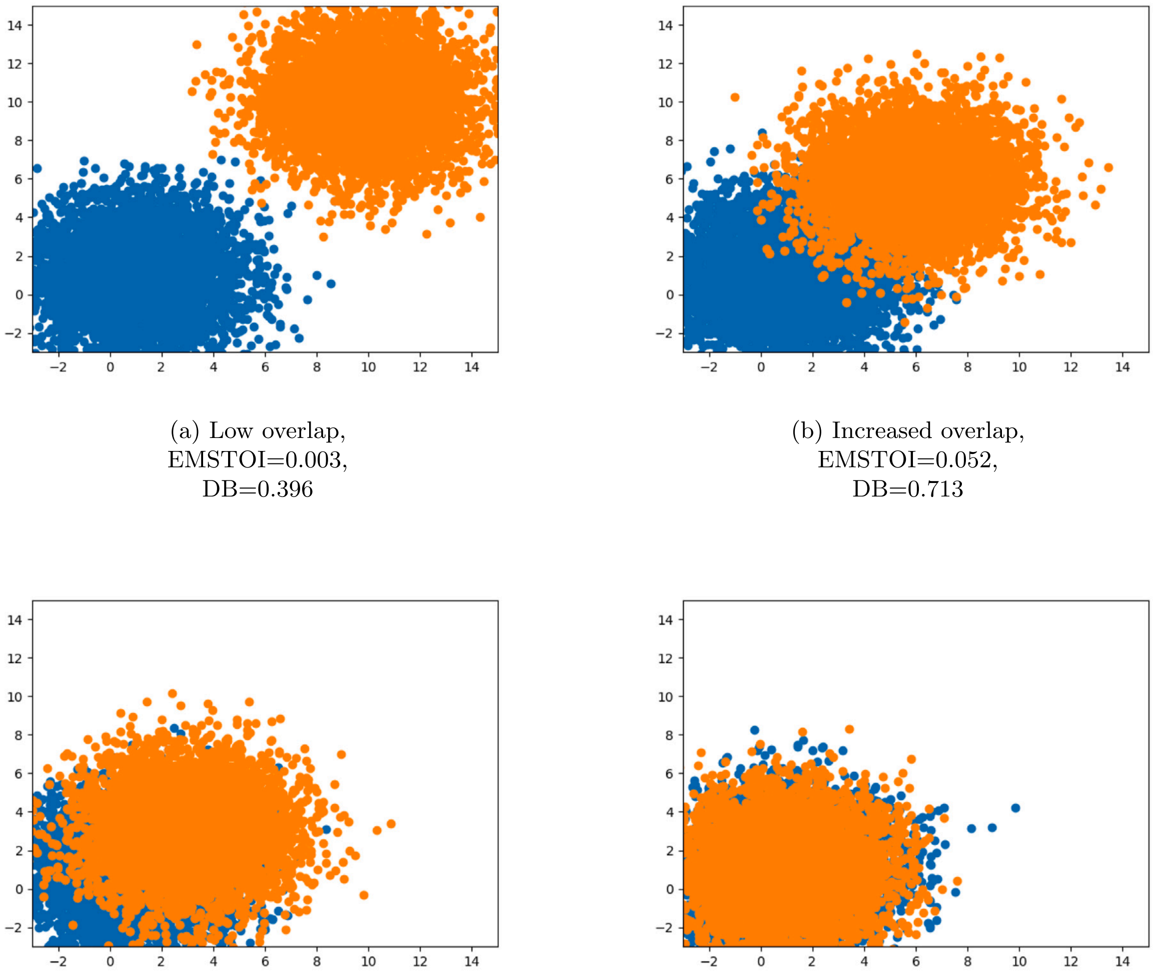
Therefore, we assess and explain the quality of an undersampled dataset by measuring its class overlap through EMSTOI and DBI. For validation, we generate several sets of undersampled datasets from a class-imbalanced dataset. We compute the classification performances across all the undersampled sets and explore the correlation between overlap indices (EMSTOI and DBI) and classification performances. Empirically, we expect a negative correlation between the overlap indices and classification performances.

3.3. Explaining and augmenting the features

We assume that there are d features and n points in a dataset, $\mathcal{D}$. We compute EMSTOI and DBI value d times, once for each feature. A feature that gives lesser EMSTOI and DBI has a better class-distinguishing ability and it plays a more instrumental role in the operation of the classifier model. For example, let there be two features f_i and f_j in $\mathcal{D}$, and their class overlap indices be $\mathcal{O}_{f_i}$ and $\mathcal{O}_{f_j}$ respectively. If $\mathcal{O}_{f_j} < \mathcal{O}_{f_i}$ and also differs by a considerable amount, we can say that feature f_j is a better distinguisher for the two classes than f_i. When the feature values are different for different classes, it can be effective in distinguishing the classes. So investigating the class overlaps in the individual features can provide explainability. However, the class overlap scores do not reveal anything about the ranges of the feature values. Without looking at the individual values of f_i and f_j for the data points, we can get an understanding of the intrinsic class overlaps using the EMSTOI and DBI values.

We further explore the correlations between these scores and the classification performances of the features. EMSTOI and DBI are low when we have well-separated classes in a dataset. A feature f_j which possesses non-overlapping ranges of values for different classes will possess low EMSTOI and DBI than another feature f_i with overlapping ranges. By virtue of the separations in the feature ranges, f_j will train an efficient classifier model, M_j (say) which can give a competent performance score. On the other hand, f_i with overlapping feature ranges for different classes is likely to output a higher overlap index. Let the classifier model trained by f_i be denoted by M_i. Consequently, f_i will train a less competent classifier model. On the same set of query points, we can expect the accuracy of M_i to be lesser than that of M_j. We will conduct our experiments on datasets from real-world domains to explore and validate this premise. We can also employ the notion of class overlap to determine the feature importance. The feature which has the lowest class overlap is the one with the highest feature importance by virtue of rendering the best possible separation of the classes. We operationalize this and compute the feature importance from the class-overlap scores which are computed through the proposed EMSTO and DBI.

We assume that there are d features in a dataset $\mathcal{D}$, denoted by $f_1, f_2, \dots, f_d$. Let $\mathcal{O}_{f_1}, \mathcal{O}_{f_2}, \dots, \mathcal{O}_{f_d}$ be class-overlap indices of $f_1, f_2, \dots, f_d$ respectively. Let us denote the importance of feature f_i, $1 \leq i \leq d$ by FI_i. We compute this value from



(a) Low overlap,
EMSTOI=0.003,
DB=0.396

(b) Increased overlap,
EMSTOI=0.052,
DB=0.713

(c) Moderate overlap,
EMSTOI=0.386,
DB=1.709

(d) High overlap,
EMSTOI=0.497,
DB=178.183

Fig. 1. A comparison of EMSTOI and DBI for quantifying the overlaps in a dataset with two classes wherein each class has exactly one cluster.

$\mathcal{O}_{f_1}, \mathcal{O}_{f_2}, \dots, \mathcal{O}_{f_d}$ as follows.

$$FI_i = \frac{\sum_{i=1}^{d} \mathcal{O}_{f_i}}{\mathcal{O}_{f_i}} \tag{3}$$

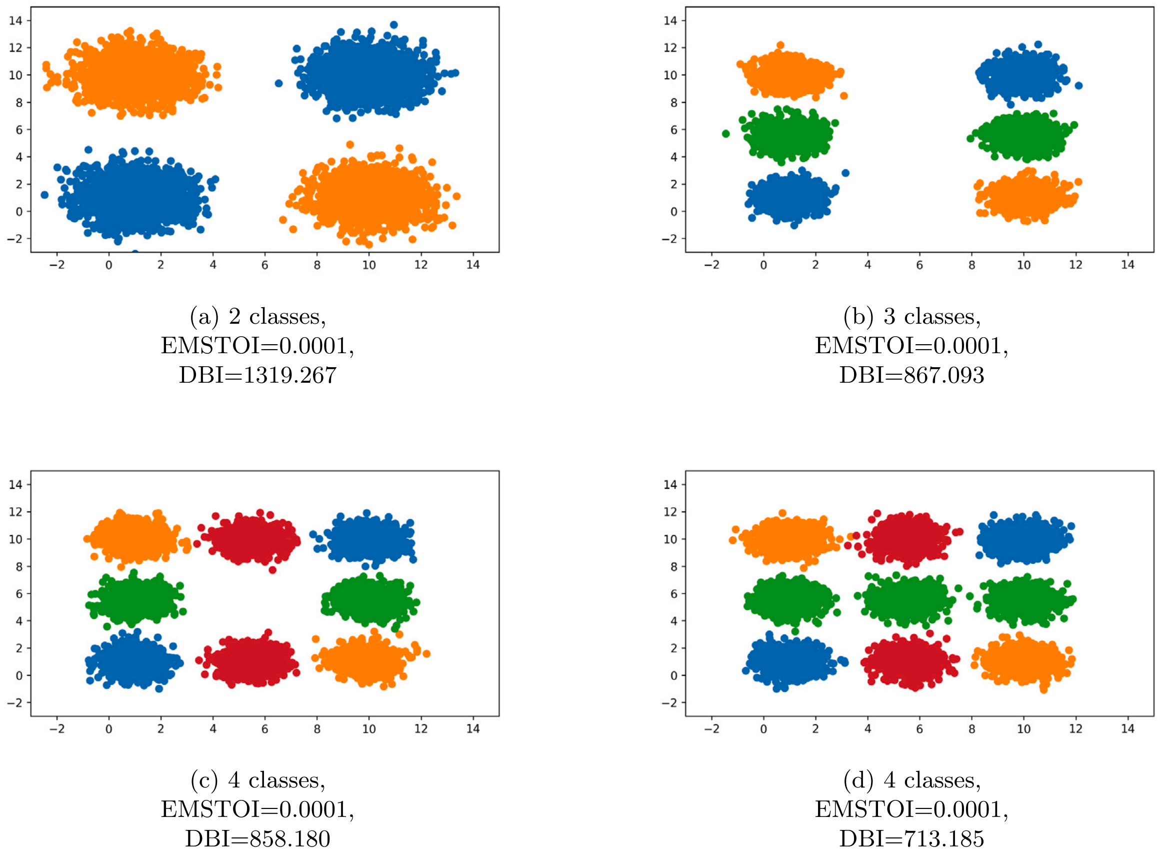
Eq. (3) signifies that FI_i is inversely proportional to the class-overlap scores, $\mathcal{O}_{f_i}$. Lesser the value of $\mathcal{O}_{f_i}$ (less overlap of the classes using that feature) of f_i, the more the feature importance. It further indicates that if the $\mathcal{O}_{f_i}$ values are equal for $1 \leq i \leq d$, the feature importance values will be equal for all the features. The computed feature importance score depends on the EMSTOI and DBI of the concerned feature as well as all other features.

4. Experimental setup

We have conducted two experimental studies to have a pre-modeling understanding of the quality of datasets (that train different models). The first study is focused on the overlap of the minority and majority classes while the second study focuses on investigation of feature importances in a dataset. We use a variety of datasets for experiments pertaining to these studies.

4.1. Datasets

For the first study, we use two synthetic datasets from the finance domain, namely *PaySim* and *BankSim*.



(a) 2 classes,
EMSTOI=0.0001,
DBI=1319.267

(b) 3 classes,
EMSTOI=0.0001,
DBI=867.093

(c) 4 classes,
EMSTOI=0.0001,
DBI=858.180

(d) 4 classes,
EMSTOI=0.0001,
DBI=713.185

Fig. 2. Comparison of EMSTOI scores and DBI in binary and multi-class synthetic datasets. In Figures (a), (b) and (c), the points belonging to a class originate from exactly two clusters (and are given a distinct color). In Figure (d), the points belonging to three classes (red, orange and blue) originate from two clusters and green originates from three clusters. In all four scenarios, the mean (resultant cluster centers) of the classes coincides in the center of the feature space. We may note that the clusters are well separated from each other, resulting in non-overlap of the classes. This is well indicated by EMSTOI which are approximately 0 in all four cases. However, DBI are high in all four cases spuriously indicating a considerable overlap in the data. This shows the robustness and advantage of EMSTOI over DBI in quantifying the overlap in the data.

PaySim dataset is constructed by simulating mobile money transactions on the basis of a sample of real transactions [59]. These are extracted from one month of financial logs from a mobile money service based in an African country. There are nine numeric and categorical features present in total in the dataset. Out of these, we have considered five numeric features namely amount of transaction, old balance of origin, a new balance of origin, old balance of destination, and new balance of destination for our study. Each transaction is categorized as *authentic* or *fraudulent*. In this dataset, 6,362,620 mobile transactions are present wherein 8213 are fraudulent thus making it a highly imbalanced dataset.

BankSim dataset is formed from an agent-based simulator of bank payments [60]. This simulation is based on a sample of aggregated transactional data provided by a bank in Spain. The main motive for constructing this synthetic data is to provide a dataset for research in the domain of fraud detection. A total of 594,643 transactions are present in this dataset, out of which, 7200 transactions are fraudulent. Hence, this dataset is also highly imbalanced. In this dataset also, there are nine numeric and categorical features present in total . We have used three numeric features age, gender of the transactor, and amount of transaction for our study. The category of each data point (transaction) is either fraudulent or normal.

For the second study, we have used two datasets originating from medical and marketing domains respectively. The first dataset namely *Maternal health* is collected from rural areas of Bangladesh [61]. It has a total of six features — age and five clinical parameters namely systolic blood pressure, diastolic blood pressure, blood sugar level, body temperature, and heart rate. Each of the 1014 datapoints consists of information on these six parameters for a pregnant woman. The women are categorized with respect to their risks for maternal mortality. Each woman belongs to any one of the following classes – *high risk, medium risk* or *low risk*.

The second dataset known as *Advertising* dataset[1] originates from the marketing domain and deals with the relationship between amounts invested in the marketing of a product through TV, radio, and newspapers (features) and the actual sales outcomes of the

[1] https://www.kaggle.com/ashydv/advertising-dataset

product. There are 200 observations in this dataset wherein the outcome of the sale is the dependent variable. Here the outcome of the sales is a continuous dependent variable and we have divided its range into five non-overlapping bins to form a multi-class classification problem.

We divide each of these datasets into two mutually exclusive and equal partitions so as to generate the training set and the test set.

4.2. Classifiers used for evaluation

We have used four classifiers across all of the experiments for evaluation and validation of the proposed approach.

- **Support Vector Classifier (SVC)** [62]: It is a supervised learning algorithm which draws a separating boundary (hyperplane) between the two classes. We have considered a linear model and allowed some tolerance for misclassification by setting the parameter $C = 1$ and $\gamma = 2$. These parameters are related to amount of misclassification allowed during training and curvature of the decision boundary, respectively.
- **k-Nearest Neighbor Classifier (KNN)** [63]: It is an intuitive and popular classification scheme. Using this classifier, the test points are classified by looking at the classes of their neighboring points. It does not involve any parameter other than the neighborhood size k. In this study, we have set $k = 1$.
- **Naive Bayes Classifier (NB)** [64]: Naive Bayes Classifier is based on the Bayes Theorem of the conditional dependence between the features and the classes of the points. A key underlying assumption in this particular classifier is the independence of the features.
- **XG-Boost Classifier** [65]: It is an ensemble method which is widely used for large datasets, because of its parallelizable nature. XGBoost consists of gradient-boosted decision trees. We have used this classifier in its default settings.

4.3. Evaluation criteria

We need two sets of criteria, one to measure the classification performance of the data, and the second to measure the correlation between the overlap indices (EMSTOI and DBI) and classification performance. For measuring the classification performance, we use *accuracy*, and *minority class* F_1. For computing the correlation between overlap indices and classification performance, we use *Pearson correlation coefficient* and *Spearman's rank correlation coefficient*. More details about these evaluation criteria are provided in Appendix A.5 and A.6.

4.3.1. Evaluation criteria for classification performance

The datasets used in this study are unbalanced. Evaluating the classification performance of such datasets is slightly different from datasets with balanced classes. Accuracy scores give an idea about the overall performance of a dataset. However, considering only the accuracy scores provide only a partial understanding of the imbalanced datasets. Therefore, we have also considered minority class F_1 to evaluate the performance of the class imbalance dataset.

4.3.2. Evaluation criteria for measuring the correlation

The main contribution of this work is the introduction of EMSTOI, through which we can measure the overlap present in the data in the pre-modeling phase itself. This knowledge is helpful in assessing the quality of the data on which a classification model is trained. We may note that DBI are also indicative of the overlap in the data to some extent however, it works well only when the classes originate from different cluster centers (Appendix A.1).

We compute the correlations between overlap indices (EMSTOI and DBI) and classifier performances. Correlation refers to the extent of the linear relationship between two variables. The value of correlation, r is $-1 \leq r \leq 1$. If two variables increase or decrease in the same direction, the sign of the correlation is positive. On the contrary, if the decrease of one leads to the increase of another and vice versa (different direction), the sign of the correlation is negative. The magnitude of the correlation value indicates the degree of relation or connection (linearity). The stronger the correlation, the greater the value, and vice versa. A correlation value of $r = 0$ indicates no correlation between the two variables. A r value close to zero (from either side) indicates a low correlation.

The two variables in our case are the performance score of the model and the overlap scores of the balanced data. If we obtain a negative correlation of a significant magnitude, we can conclude that our approach is viable. We compute the correlations between performance and overlap scores via two standard metrics *Pearson correlation coefficient* and *Spearman rank correlation coefficient*. The details of these coefficients is provided in Appendix A.6.

5. Results and analysis

In this section, we report the results of the two sets of experiments. In the first study, we utilize the *PaySim* and the *BankSim* datasets. In this study, we investigate the correlation between overlap indices (EMSTOI and DBI) and classification performance followed by the association between overlap indices and the amount of money saved and money lost. Subsequently, the second study is conducted using *Maternal Health* and *Advertising* datasets where our objective is to ascertain the feature importance through overlap indices.

Table 1
Results of correlation of *EMSTOI* and classification performance on *PaySim* and *BankSim* datasets.

	PaySim				BankSim			
Classifier	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
	F_1		Accuracy		F_1		Accuracy	
SVC	-0.226	-0.214	-0.255	-0.177	-0.283	-0.242	-0.265	-0.188
KNN	-0.309	-0.287	-0.230	-0.192	-0.251	-0.284	-0.287	-0.269
NB	-0.252	-0.279	-0.301	-0.386	-0.188	-0.257	-0.287	-0.288
XGBoost	-0.179	-0.223	-0.400	-0.389	-0.292	-0.273	-0.202	-0.168

Table 2
Results of correlation between *DBI* and classification performance on *PaySim* and *BankSim* datasets.

	PaySim				BankSim			
Classifier	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
	F_1		Accuracy		F_1		Accuracy	
SVC	-0.222	-0.206	-0.227	-0.236	-0.271	-0.142	-0.149	-0.051
KNN	-0.247	-0.252	-0.071	-0.127	-0.165	-0.166	-0.138	-0.127
NB	-0.263	-0.217	-0.223	-0.192	-0.262	-0.217	-0.170	-0.156
XGBoost	-0.267	-0.110	-0.235	-0.174	-0.191	-0.169	-0.134	-0.122

Table 3
Results of correlation between *EMSTOI* and amount lost and amount saved on *PaySim* and *BankSim* datasets.

	PaySim				BankSim			
Classifier	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
	Amount lost		Amount saved		Amount lost		Amount saved	
SVC	0.281	0.196	-0.323	-0.282	0.264	0.139	-0.207	-0.183
KNN	0.270	0.265	-0.222	-0.175	0.195	0.236	-0.278	-0.315
NB	0.184	0.215	-0.404	-0.328	0.396	0.257	-0.235	-0.348
XGBoost	0.205	0.245	-0.434	-0.361	0.243	0.132	-0.199	-0.251

Table 4
Results of correlation between *DBI* and amount lost, and amount saved on *PaySim* and *BankSim* datasets.

	PaySim				BankSim			
Classifier	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
	Amount lost		Amount saved		Amount lost		Amount saved	
SVC	0.092	0.152	-0.204	0.186	0.143	0.071	-0.212	-0.133
KNN	0.115	0.116	-0.225	-0.249	0.171	0.180	-0.262	-0.317
NB	0.159	0.164	-0.291	-0.262	0.082	0.160	-0.213	-0.163
XGBoost	0.133	0.206	-0.112	-0.133	0.224	0.174	-0.213	-0.221

5.1. Analyzing the correlation between overlap indices and classification performance

The objective of the first study is to investigate the correlation between class overlap and classification performance. The class overlaps in the undersampled datasets are measured through EMSTOI and DBI. We also evaluate the classification performance of different classifiers trained on the undersampled *PaySim* and *BankSim* datasets. We report the correlation results in Table 1 and Table 2.

For each dataset, there are four classifiers (SVC, KNN, NB, and XGBoost), two performance evaluation metrics (accuracy and F_1), and two correlation metrics (Person and Spearman) — leading to $(4 \times 2 \times 2 =)$ 16 cases. For both *PaySim* and *BankSim* datasets, we have repeated this process 50 times (obtaining 50 sets of balanced datasets) and computed the correlation between classification performance (accuracy and minority class F_1) and the overlap score (EMSTOI and DBI). As shown in Table 1, for both datasets, a weak negative correlation (value ≤ -0.2) is obtained between EMSTOI and classification performance in 13 out of 16 cases. These results are in congruence with our expectations and it can be deduced that classification performances could be explained through the EMSTOI scores. In terms of DBI (Table 2), a weak negative correlation [66] is obtained between DBI and classification performance 11 out of 16 times on *PaySim* dataset and 3 out of 16 times on *BankSim* dataset. These results demonstrate the superiority of the EMSTOI in assessing the quality of datasets on which different classifiers are trained.

5.2. Analyzing the correlation between overlap indices and the amount of money saved and lost

We investigate the correlation between overlap indices (EMSTOI and DBI) and the amount saved through the detection of fraudulent transactions. We also investigate the correlation between overlap indices (EMSTOI and DBI) and the amount lost by missing the detection of fraudulent transactions. The results for EMSTOI and DBI are reported in Table 3 and Table 4, respectively.

A lesser overlap between the minority class and the majority class of a balanced dataset indicates quality data. This quality data is supposed to learn an efficient classifier model which can detect fraudulent transactions accurately and save more amount of money.

Table 5
Results of feature-specific overlap indices (EMSTOI) and accuracy of *Maternal Health* dataset.

Features	Overlap Indices		Classifiers			
	EMSTOI	DBI	SVC	KNN	NB	XG-Boost
Age	0.624	8.116	0.484	0.519	0.479	0.502
Systolic Blood Pressure	0.529	3.511	0.590	0.636	0.618	0.641
Diastolic Blood Pressure	0.609	4.920	0.552	0.558	0.581	0.581
Blood Sugar	0.486	2.557	**0.622**	**0.647**	**0.645**	**0.668**
Body Temperature	0.633	6.375	0.535	0.539	0.544	0.525
Heart Rate	0.687	15.248	0.464	0.502	0.539	0.567

Table 6
Results of feature-specific overlap indices (EMSTOI) and accuracy of *Advertising* dataset.

Features	Overlap Indices		Classifiers			
	EMSTOI	DBI	SVC	KNN	NB	XG-Boost
TV	0.452	3.521	**0.540**	**0.540**	**0.500**	**0.500**
Radio	0.681	22.418	0.480	0.350	0.340	0.340
Newspaper	0.772	29.283	0.320	0.230	0.280	0.280

Table 7
Classification accuracy obtained on original data, data weighted by *EMSTOI* driven feature importance, and data weighted by *DBI* driven feature importance.

	Classifiers			
	SVC	KNN	NB	XG-Boost
Maternal Health				
Original data	0.660	0.582	0.635	0.624
Data weighted by FI-EMSTOI	0.571	**0.609**	**0.651**	**0.632**
Data weighted by FI-DBI	0.566	**0.614**	**0.643**	**0.637**
Advertising				
Original data	0.708	0.740	0.640	0.747
Data weighted by FI-EMSTOI	**0.806**	**0.797**	**0.650**	**0.768**
Data weighted by FI-DBI	**0.736**	**0.755**	**0.648**	**0.762**

On the other hand, an increased overlap of the different classes will train a sub-optimal classifier. Such a model will misclassify the fraudulent transactions and increase the loss incurred. We have looked into these perspectives and computed the average amount saved and lost through the detection of fraudulent transactions and the overlap indices of the balanced datasets. For both *PaySim* and *BankSim* datasets, we have repeated this process 50 times (obtaining 50 sets of balanced datasets) and computed the correlation between the amount saved/lost and the overlap indices. The correlations with the EMSTOI are reported in Table 3 and with DBI are reported in Table 4. As expected both EMSTOI and DBI are positively correlated with the amount lost and negatively correlated with the amount saved on both datasets. It is noteworthy that for the correlation with the amount lost, EMSTOI outperforms DBI (6/8 vs. 1/8) on *PaySim* dataset as well as (5/8 vs. 1/8) on *BankSim* dataset. Similarly, for correlation with the amount saved, EMSTOI performs at par with DBI (7/8 vs. 5/8) on *PaySim* dataset and (6/8 vs. 6/8) on *BankSim* dataset. These results substantiate that EMSTOI can better assess the quality of data used for training a classification model.

5.3. Analyzing the feature importance using overlap indices

We also study the association between class overlap and classification performance for each individual feature. We report the class overlap indices (EMSTOI and DBI) and the accuracy values obtained through four classifiers (SVM, KNN, NB, and XGBoost) for all the 6 features of *Maternal health* and 3 features of *Advertising* dataset in Table 5 and Table 6, respectively. These results show a strong association between the class-overlap indices of the features and the accuracy of classification.

On *Maternal health* dataset, the lowest EMSTOI and DBI are obtained for *Blood Sugar*, and the best classification performance is delivered by this feature on all four classifiers (Table 5). Additionally, in the *Advertising* dataset, the lowest EMSTOI and DBI are obtained for *TV*, and the best classification performance is delivered by this feature on all four classifiers (Table 6). These findings establish the utility of the proposed approach in determining the feature importance by examining the association between feature overlap scores (EMSTOI and DBI) and their respective classification performances. It is important to note that the feature importances obtained by our approach are in the pre-modeling phase and without the intervention of the test phase. Our approach is also model-agnostic and depends only on the training data.

Further, we incorporate the feature importance scores in the data before building the model. Prior to training the model, we created feature importance-informed training data. This was accomplished by adding weights to the features in decreasing order of their EMSTOI. We investigated the outcome of this augmentation on classification performance. We demonstrate this through experiments on *Maternal Health dataset* and *Advertising* datasets. For these experiments, half of the data points are used in model building, and the remaining half is used in prediction. The results of over 100 independent runs on four classifiers are reported in Table 7. It shows that the use of weighted features has offered better performance on three out of four classifiers on *Maternal Health dataset*. On *Advertisement* dataset, the use of weighted features has improved the performances across all four classifiers.

6. Discussion

In this study, we posit that the quality of data is integral to the supervised learning tasks and their business or financial outcomes. We consider the problem of explaining the performance of a classifier in the pre-modeling phase by assessing the quality of data. The proposed index EMSTOI offers an advantage in terms of data induced, model-agnostic understanding of classification performance. The proposed Euclidean minimum spanning tree-based index helps in appropriately measuring class overlaps in data wherein the classes can originate from single or multiple clusters. It is noteworthy that when classes originate from multiple clusters, state-of-the-art indices such as DBI may fail to accurately assess the class-overlap. Experiments and comparative analysis using datasets from a variety of domains demonstrate that the proposed approach yields superior assessment of training data. We transform the data-points into a connected graph by constructing their Euclidean Minimum Spanning Tree. Subsequently, the proposed index captures the ratio of heterogeneous edges over homogeneous edges in an EMST transformed dataset, and the ratio of average weight of homogeneous edges over the average weight of heterogeneous edges thereby helping in quantifying the class overlaps in data wherein classes may originate from multiple clusters.

The results of classification are not only based on the characteristics of the model but also on the data which is used to learn the model. Therefore, the examination of data can provide insights in explaining the output of a model. Our experimental analysis shows that there is indeed a correlation between class overlap and classification performance wherein EMSTOI is better correlated (as compared to DBI) with classification accuracy as well as minority class F1 scores. EMSTOI also correlates better with the amount saved or lost by detecting fraudulent transactions in terms of both correlation coefficients namely Pearson correlation coefficient and Spearman rank correlation coefficient. Further, the overlap indices can also be used to derive the importance of different features and classification performance based on each feature. Thus using data from diverse domains, we analyze and demonstrate the viability of the proposed approach. The use of EMSTOI in comparison to DBI in assessing the quality of data which is used to train classification models represents the key methodological contribution of this study.

This study responds to recent call for research to find innovative ways to enhance and understand classification performance thereby helping in the more effective exploration of the search universe [67]. It is important to note that several classification models have been developed and are in use among researchers and practitioners [62–65]. These models demonstrate a wide variety of capabilities depending on the domain for which they are being used. Through the experimental analysis, we demonstrate that the proposed approach for explaining the classification performance is model as well as domain agnostic. The proposed index will provide an impetus to research in the field of explainable AI for knowledge creation.

6.1. Theoretical implications

This study has several theoretical implications. It helps in advancing understanding of how the different features and their meta-characteristics are related to each other, given a dataset related to a particular domain. This study also suggests that classification performance can be significantly improved by investigating data in the pre-modeling phase, rather than doing one post hoc validation at the end. Such middle-range theorizing is crucial to advance the literature on how emerging AI and ML models can deliver superior business value [68]. Deducing explainability through interpretable parameters of the data is a natural progression towards advancements in AI modeling [69].

The proposed approach offers scholars a new way to reason about empirically observed variations in the classification performance of models. The introduction of Euclidean Minimum Spanning Tree based approach also offers a systematic way to analyze the geometric properties of data thereby suggesting the need to revisit existing data quality metrics by adopting graph-theoretic principles [70]. This approach can overcome challenges arising in a variety of business environments by allowing stakeholders to gain insights into model behavior without compromising the confidentiality of data.

Also, the experimental results demonstrate that the proposed index can discern feature importance more effectively in multi-class datasets. This advances theoretical understanding of how feature spaces can be optimized for improved classifier performance, thereby offering new feature selection strategies which prioritize intrinsic class separability [71]. The experimental validation across diverse domains, including finance, medical, and marketing datasets, demonstrate the generalizability of the proposed index for data quality assessment. Future studies can expand upon theoretical implications of this study by exploring additional graph-theoretic artifacts or applying the proposed index in novel contexts such as unsupervised learning or data augmentation. We hope that the insights of this study inspire future research designs to assess the quality of collected data prior to empirically analyzing it using different types of quantitative models.

6.2. Managerial implications

In today's digital environment, there is a lot of heterogeneity in the data generated from a variety of applications [72]. The business solutions that classification models offer have gained traction as they simplify mundane tasks while enabling organizations to achieve productive results [73]. Considering such a scenario, classification models need to accommodate the data with varying structure, characteristics, and quality in an effective way. At the same time, enterprises may like to be able to use interpretable measures of data quality wherein they can ascertain the appropriate classification model for their work. As such, managers can use the proposed index to evaluate the readiness of datasets for classification tasks, ensuring that classification models are built on reliable data.

The increasing expectations of customers and the pursuit to improve classification performance are driving researchers to develop new classification approaches without considering the intrinsic characteristics of data [55]. The proposed approach can not only generate significant gains for organizations by offering explainability of classification tasks but also allowing customization of classification models that serve the needs of diverse sets of users. This can help build trust among stakeholders, by demonstrating the logical foundations of model decisions without compromising data security. This is particularly critical for industries dealing with sensitive information, such as healthcare and finance, where compliance with regulations like GDPR or HIPAA is mandatory [74].

The ability of the proposed EMSTOI to determine feature importance in multi-class datasets can guide managers in prioritizing data attributes that matter most for predictive accuracy. This targeted focus can streamline data collection processes. Moreover, the proposed index can be used by managers to identify classification challenges early in the machine learning pipeline. Overall, the proposed approach has the potential to minimize the risk of deploying underperforming models, reducing financial losses and reputational damage associated with misclassifications.

6.3. Limitations and future research directions

This study is not devoid of limitations. First, the generalizability of the results to unstructured data, such as image, time series, networked data or text datasets remains to be established. Second, the computational complexity for constructing and computing EMSTOI may pose challenges when applied to large-scale datasets with high dimensionality. Third, the reliance on Euclidean distance assumes that data distributions are well represented in Euclidean space. This may limit the method's effectiveness for data where relationships are better captured by non-Euclidean distance measures [75] Finally, the offline experimentation performed in this study cannot completely replace online real-time experimentation. As a result, the offline laboratory experimental approach used in this work has inherent limitations. Nonetheless, we believe that this study will guide future research and add rigor to the field of data quality assessment for AI modeling.

Future research studies in the area of data quality assessment can focus on exploring how the proposed approach performs for unsupervised, semi-supervised, and reinforcement learning environments. Secondly, future studies could focus on integrating the proposed index with automated feature selection algorithms to optimize dataset quality and improve model training. Thirdly, since we use only DBI as a benchmark in this study, future studies should validate the proposed index against a wider array of existing measures, such as Silhouette Score, and Calinski–Harabasz Index [76,77] Finally, the experimental datasets used in this study are completely labeled. In the future, an augmented approach can be developed which will work well when the data is partially labeled. We hope that future work following the aforementioned directions would concentrate on additional in-depth aspects of the proposed approach in a more granular way.

7. Conclusion

In this study, we present a graph-theoretic approach for assessing data quality in classification tasks. This study addresses a critical need for effective measures that connect model performance with data characteristics. We propose a new index for quantifying the intrinsic separations between classes in data. Through experiments on datasets from finance, medical, and marketing domains, we validated the effectiveness of the proposed index against the widely used Davies–Bouldin Index (DBI). The proposed index enables determination of feature importance and exhibits superior correlation with classification performance thereby demonstrating its utility in multi-class datasets. These findings underscore the practical relevance of the measure, offering new avenues for data quality assessment, enhancing classifier design and performance evaluation. This study contributes in two important ways. First, it establishes a reliable and generalizable measure for data quality assessment. Second, by demonstrating the potential of proposed approach to advance classification tasks across diverse domains. Our empirical studies show that the insights on the class separations can be used to shed light on several important aspects such as the amount of money saved by detecting fraudulent transactions, the amount of money lost by missing the detection of fraudulent transactions, and learning the feature importance. A key advantage of the proposed approach pertains to its application in the pre-modeling phase which can offer explanations about a model's performance without even building the model. This study enriches the toolkit for data quality assessment and contributes towards building more effective AI systems.

CRediT authorship contribution statement

Payel Sadhukhan: Writing – original draft, Methodology, Investigation, Conceptualization. **Samrat Gupta:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Ethics approval

Not Applicable. This article does not contain any studies with human participants or animals performed by any of the authors.

Declaration of Generative AI and AI-assisted technologies in the writing process

Not used.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The corresponding author (Samrat Gupta) is supported by a funded project on FAME data marketplace under European Union’s Horizon research and innovation programme (Grant Agreement No. : 101092639).

Appendix A

A.1. Davies–bouldin index

Let us have a set of points in $\mathcal{R}^n$ and let C_i be a cluster of datapoints. We consider $\mathbf{x}_j$ to be a datapoint belonging to cluster C_i. We assume that we will carry out the computations in Euclidean space. We denote the n-dimensional cluster center of C_i with A_i and the cardinality of C_i (number of points in C_i) to be T_i. The within cluster distance between the points of a cluster, C_i is denoted with S_i.

$$S_i = \frac{1}{T_i}\left(\sum_{j=1}^{T_i} \|\mathbf{x}_j - A_j\|^2\right)^{\frac{1}{2}} \tag{4}$$

We denote the separation between clusters C_i and C_j with $M_{i,j}$.

$$M_{i,j} = \| \left(A_i - A_j\right)^2 \|^{\frac{1}{2}} \tag{5}$$

Let $R_{i,j}$ be a measure of evaluating the vitality of clustering scheme. This measure will depend on two things —- [i] S_i, within cluster distance of C_i for all the clusters — this has to be as low as possible and, [ii] $M_{i,j}$, the separation between two different clusters C_i and C_j — this has to be as high as possible.

$R_{i,j}$ is defined in terms of S_i and $M_{i,j}$ as follows.

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}} \tag{6}$$

This formulation satisfies the following conditions:

[i] $R_{i,j} \geq 0$.

[ii] $R_{i,j} = R_{j,i}$.

[iii] When $S_j \geq S_k$ and $M_{i,j} = M_{i,k}$, we get $R_{i,j} \geq R_{i,k}$.

[iv] When $S_j = S_k$ and $M_{i,j} \leq M_{i,k}$, we get $R_{i,j} \geq R_{i,k}$.

Lower the value of $R_{i,j}$, better is the separation between clusters C_i and C_j. On the contrary, a higher value of $R_{i,j}$ indicates increased similarity between the clusters. When the number of clusters obtained in a dataset is more than 2, the highest similarity score for a cluster (with respect to another cluster) is taken into account.

$$D_i = \max_{j \neq i} \; R_{i,j} \tag{7}$$

Davies–Bouldin Index (DBI) is the cumulative sum of D_i over all clusters in a given dataset. It shows that the overall similarity or overlap of the clusters. Lower the value of DB, more separated are the clusters. For a dataset with N clusters, DB is defined as follows.

$$DBI = \frac{1}{N}\sum\nolimits_{i=1}^{N} \; D_i \tag{8}$$

The correctness of DBI in rendering class overlap is dependent on a stringent assumption stated as follows. The cluster centers have to be distinct from each other and should possess a substantial amount of separation. If the separation is less, DBI will not be indicative of the overlap present in the data. When the classes originate from different cluster centers, DB works well in measuring the overlap between them.

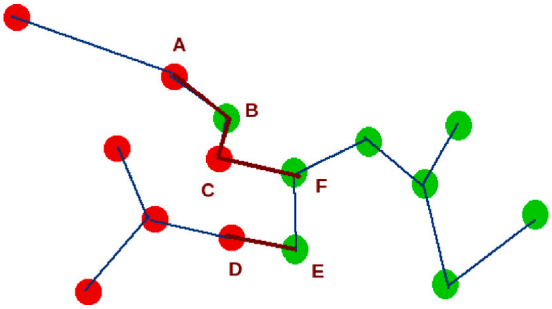


Fig. 3. This figure shows the minimum spanning tree (MST) for 15 arbitrary points. The datapoints belong to class 1 (indicated by red) or class 2 (indicated by green). When the two vertices of an edge belong to the same class, it is termed a *homogeneous edge* (the blue colored edges in the given figure). On the other hand, when the two vertices on edge belong to two different classes, we term it a *heterogeneous edge* (AB, BC, CF, and ED in the given figure.).

A.2. Euclidean minimum spanning tree

The Euclidean minimum spanning tree (EMST) is a minimum spanning tree (MST) of a set of n points in $\mathcal{R}^d$, where the weight of an edge between each pair of points is the *euclidean distance* between the two points. An EMST connects a set of points using edges such that the total weight of the edges is minimized and each point is reachable from any other point through the edges.

For EMST construction, it is assumed that we have a complete graph for a set of n points. The edge weight between any two points is their euclidean distance. Hence, we have a graph $G(V, E)$ where V is the vertex set, the set of points in $\mathcal{R}^d$ and E denotes the edge set. EMST is a sub-graph H of G ($H \subset G$) in terms of edges satisfying the following two conditions. Let e_H be the total weight of edges in H.

[i] A vertex v of H is reachable from another vertex u of H, $\forall u, v \in H$.

[ii] We satisfy condition (i) and the sum of edge weights of H, e_H is minimum.

A.3. Mathematical foundation of the EMST-based class-overlap index

Once we form the EMST from n points in a feature space, we have a connected graph of n vertices. A key characteristic of EMST (and also a MST) is — the points are connected in the most compact fashion by virtue of minimizing the total edge weights. The connection weight (or the edge-weight) between any two points indicate their proximity. The task that we are addressing in this work is quantifying the degree of class-overlap in a dataset. Two aspects of an EMST are particularly interesting and can serve as a data mine in this regard, they are — [i] the class of the two vertices connecting an edge, and, [ii] the weight of that edge. Intuitively, in a dataset with well separated classes (low overlap), the edges of an EMST will be mostly formed between vertices belonging to the same class, which are termed as *homogeneous edges*. An edge whose two vertices belong to different classes is known as *heterogeneous edge* (Fig. 3). Since EMST is a connected graph, we will indeed have at least one *heterogeneous edge*.

A.4. Mathematical properties of EMSTOI

EMSTO renders an overlap value between 0 and ∞ for a dataset. Such an extended and unlimited range can come in the way of proper comprehension of overlap existing in a dataset, and while comparing the class overlaps in two different datasets. Motivated to address this concern, we tweaked the calculation to restrict the overlap quantification in $[0, 1]$ in the following manner.

$$\mathrm{EMSTOI} = \frac{\mathrm{EMSTO}}{\mathrm{EMSTO} + 1} \tag{9}$$

The function $f(x) = \frac{x}{x+1}$ is strictly increasing as $f(x_1) < f(x_2) \quad \forall x_1 < x_2$ in the domain of f. The strictly increasing property can be verified by obtaining its derivative and applying the quotient rule in the following manner.

We compute its derivative as follows:

$$f'(x) = \frac{d}{dx}\left(\frac{x}{x+1}\right)$$

Applying the quotient rule, $\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{u'v-uv'}{v^2}$, where $u = x$ and $v = x + 1$:

$$f'(x) = \frac{(1)(x+1) - (x)(1)}{(x+1)^2} = \frac{x+1-x}{(x+1)^2} = \frac{1}{(x+1)^2}$$

We may note that,

1. $f'(x) = \frac{1}{(x+1)^2}$ is positive for all $x > -1$, including $[0, \infty)$.

2. A positive derivative affirms strictly increasing nature of the function.

Hence, the order of EMSTO values is preserved in EMSTOI. The least value of EMSTOI (0) is obtained when EMSTO is 0. When EMSTO=∞, we get EMSTOI value of 1 which can be deduced through L'Hospital's Rule [78].

A.5. Definition of the evaluating metrics

- **Accuracy** [79] score computes the fraction of correctly classified instances (transactions). The higher the accuracy score, better is the classification performance, the computed value ranges from 0 and 1 (both inclusive). We define *accuracy* as follows:

$$accuracy = \frac{\text{number of correct predictions}}{\text{total number of points}} \tag{10}$$

- **Minority class $\mathbf{F_1}$** [79] is dependent on the correctness of the prediction of the datapoints belonging to the minority class (fraudulent transactions). Let N be the total number of data points and let the points in the dataset belong to only one of the two classes, class 1 (minority class) and class 0 (majority class). Let *True Positive (TP)* denote the number of datapoints correctly classified to class 1 (minority class, fraudulent class). Similarly, *True Negative (TN)* denotes the number of datapoints correctly classified to class 0 (majority class, authentic transactions). *False Positive (FP)* denotes the number of datapoints belonging to class 0 (negative or majority) but have been classified as class 1 (positive or minority) by the model (predicted as fraudulent, but actually authentic). Similarly, *False Negative (FN)* denotes the number of datapoints belonging to class 1 (positive or minority) but have been classified as class 0 (negative or majority) by the model (predicted as authentic, but actually fraudulent).
Hence, $N = TP + TN + FP + FN$
Precision (for the minority or the fraudulent class) is the number of points correctly classified as a minority (fraudulent transaction) scaled by the total number of minority predictions (including the predictions misclassified as fraudulent).

$$precision = \frac{TP}{TP + FP} \tag{11}$$

Recall (for the minority or the fraudulent class) is the number of points correctly classified as a minority (fraudulent transaction) scaled by the total number of fraudulent transactions present in the data (including the predictions misclassified as authentic).

$$recall = \frac{TP}{TP + FN} \tag{12}$$

Minority class F_1 is the harmonic mean of the *precision* and *recall* for the minority class. It indicates the fidelity of the classifier's decision in the context of the minority class (fraudulent class in this experiment). The higher the value of this metric, the better the performance. The computed value lies between 0 and 1.

$$\text{Minority class } F_1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{13}$$

A.6. Definition of the metrics assessing the correlation

- **Pearson correlation coefficient, P(X,Y)** [79]: Technically, it is the ratio of the covariance of the two variables and the product of the variances of the variables. Let X and Y be the two variables. Let σ_X and σ_Y denote the variances of X and Y respectively. We denote the covariance of X and Y with $cov(X, Y)$.

$$\text{P(X,Y)} = \frac{cov(X, Y)}{\sigma_X * \sigma_Y} \tag{14}$$

- **Spearman's rank correlation coefficient, SP(X,Y)** [79]: In essence, it is similar to the previous metric. However, the key difference is, instead of considering the values, Spearman's rank correlation coefficient works on the rank of the values. Let X and Y be the two variables for which we want to compute the correlation. Let $R(X)$ and $R(Y)$ be the ranks of the items of X and Y. Let $\sigma_{R(X)}$ and $\sigma_{R(Y)}$ denote the variances of $R(X)$ and $R(Y)$ respectively. We denote the covariance of $R(X)$ and $R(Y)$ with $cov(R(X), R(Y))$.

$$\text{SP(X,Y)} = \frac{cov(R(X), R(Y))}{\sigma_{R(X)} * \sigma_{R(Y)}} \tag{15}$$

Data availability

Data will be made available on request.

References

[1] J.-C. Pomerol, Artificial intelligence and human decision making, European J. Oper. Res. 99 (1) (1997).
[2] G. George, E.C. Osinga, D. Lavie, B.A. Scott, Big data and data science methods for management research, Acad. Manag. J. 59 (5) (2016) 1493–1507.
[3] B.C. Stahl, A. Andreou, P. Brey, T. Hatzakis, A. Kirichenko, K. Macnish, S.L. Shaelou, A. Patel, M. Ryan, D. Wright, Artificial intelligence for human flourishing–beyond principles for machine learning, J. Bus. Res. 124 (2021).
[4] W. Samek, K.-R. Müller, Towards explainable artificial intelligence, in: Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Springer, 2019, pp. 5–22.

[5] G. Smith, Data mining fool's gold, J. Inf. Technol. 35 (3) (2020) 182–194.

[6] A. Rai, Explainable AI: from black box to glass box, J. Acad. Mark. Sci. 48 (1) (2020) http://dx.doi.org/10.1007/s11747-019-00710-5.

[7] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (5) (2019) 206–215.

[8] M.S. Santos, P.H. Abreu, N. Japkowicz, A. Fernández, C. Soares, S. Wilk, J. Santos, On the joint-effect of class imbalance and overlap: a critical review, Artif. Intell. Rev. 55 (8) (2022) 6207–6275.

[9] S. Petrovic, A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters, in: Proceedings of the 11th Nordic Workshop of Secure IT Systems, vol. 2006, Citeseer, 2006, pp. 53–64.

[10] S. Maldonado, C. Vairetti, A. Fernandez, F. Herrera, FW-SMOTE: A feature-weighted oversampling approach for imbalanced classification, Pattern Recognit. 124 (2022) 108511, http://dx.doi.org/10.1016/j.patcog.2021.108511.

[11] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, G. Bing, Learning from class-imbalanced data: Review of methods and applications, Expert Syst. Appl. 73 (2017) 220–239.

[12] M. Spiekermann, Data marketplaces: Trends and monetisation of data goods, Intereconomics 54 (4) (2019) 208–216.

[13] M. Zhang, F. Beltrán, J. Liu, A survey of data pricing for data marketplaces, IEEE Trans. Big Data 9 (4) (2023) 1038–1056.

[14] M.T. Ribeiro, S. Singh, C. Guestrin, "Why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[15] M. Moradi, M. Samwald, Post-hoc explanation of black-box classifiers using confident itemsets, Expert Syst. Appl. 165 (2021) 113941.

[16] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[17] B. Dushimimana, Y. Wambui, T. Lubega, P.E. McSharry, Use of machine learning techniques to create a credit score model for airtime loans, J. Risk Financ. Manag. 13 (8) (2020) 180.

[18] C. Sánchez, S. Maldonado, C. Vairetti, Improving debt collection via contact center information: A predictive analytics framework, Decis. Support Syst. 159 (2022) 113812, http://dx.doi.org/10.1016/j.dss.2022.113812.

[19] O. Ore, M. Sposato, Opportunities and risks of artificial intelligence in recruitment and selection, Int. J. Organ. Anal. 30 (6) (2022) 1771–1782.

[20] K. Topuz, B. Davazdahemami, D. Delen, A Bayesian belief network-based analytics methodology for early-stage risk detection of novel diseases, Ann. Oper. Res. (2023) http://dx.doi.org/10.1007/s10479-023-05377-4.

[21] A. Ghai, P. Kumar, S. Gupta, A deep-learning-based image forgery detection framework for controlling the spread of misinformation, Inf. Technol. People 37 (2) (2024) 966–997.

[22] E.D. Zamani, C. Smyth, S. Gupta, D. Dennehy, Artificial intelligence and big data analytics for supply chain resilience: a systematic literature review, Ann. Oper. Res. 327 (2) (2023) 605–632.

[23] X. Li, V. Krivtsov, C. Pan, A. Nassehi, R.X. Gao, D. Ivanov, End-to-end supply chain resilience management using deep learning, survival analysis, and explainable artificial intelligence, Int. J. Prod. Res. (2024) 1–29.

[24] S. Gupta, S. Kumar, P. Kumar, Evaluating the predictive power of an ensemble model for economic success of Indian movies, J. Predict. Mark. 10 (1) (2016) 30–52.

[25] Á. Delgado-Panadero, B. Hernández-Lorca, M.T. García-Ordás, J.A. Benítez-Andrades, Implementing local-explainability in gradient boosting trees: Feature contribution, Inform. Sci. 589 (2022).

[26] J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955, AI Mag. 27 (4) (2006) 12–12.

[27] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.

[28] V.N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 10 (5) (1999) 988–999.

[29] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[30] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[31] Z.C. Lipton, The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery, Queue 16 (3) (2018) 31–57.

[32] T. San Kim, S.Y. Sohn, Machine-learning-based deep semantic analysis approach for forecasting new technology convergence, Technol. Forecast. Soc. Change 157 (2020) 120095.

[33] C. DeBrusk, The risk of machine-learning bias (and how to prevent it), MIT Sloan Manag. Rev. (2018).

[34] A. Adadi, M. Berrada, Peeking inside the black-box: a survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018).

[35] G. Yang, Q. Ye, J. Xia, Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond, Inf. Fusion 77 (2022).

[36] G. Vilone, L. Longo, A quantitative evaluation of global, rule-based explanations of post-hoc, model agnostic methods, Front. Artif. Intell. 4 (2021) 717899.

[37] R. Confalonieri, T. Weyde, T.R. Besold, F.M. del Prado Martín, Using ontologies to enhance human understandability of global post-hoc explanations of black-box models, Artificial Intelligence 296 (2021) 103471.

[38] M.T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, (1) 2018.

[39] C. Nóbrega, L. Marinho, Towards explaining recommendations through local surrogate models, in: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, 2019, pp. 1671–1678.

[40] G. Ten Broeke, G. van Voorn, A. Ligtenberg, J. Molenaar, The use of surrogate models to analyse agent-based models, J. Artif. Soc. Soc. Simul. 24 (2) (2021).

[41] B.M. Greenwell, pdp: an R package for constructing partial dependence plots, R J. 9 (1) (2017) 421.

[42] D.W. Apley, J. Zhu, Visualizing the effects of predictor variables in black box supervised learning models, 2016, arXiv preprint arXiv:1612.08468.

[43] A. Goldstein, A. Kapelner, J. Bleich, E. Pitkin, Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation, J. Comput. Graph. Statist. 24 (1) (2015) 44–65.

[44] A. Hassan, J.H. Paik, S. Khare, S.A. Hassan, PPFS: Predictive permutation feature selection, 2021, arXiv preprint arXiv:2110.10713.

[45] S. Wachter, B. Mittelstadt, C. Russell, Counterfactual explanations without opening the black box: Automated decisions and the GDPR, Harv. JL Tech. 31 (2017).

[46] A.-H. Karimi, G. Barthe, B. Balle, I. Valera, Model-agnostic counterfactual explanations for consequential decisions, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2020.

[47] L. Geng, H.J. Hamilton, Choosing the right lens: Finding what is interesting in data mining, in: Quality Measures in Data Mining, Springer, 2007, pp. 3–24.

[48] J.-M. Gaillard, M. Hebblewhite, A. Loison, M. Fuller, R. Powell, M. Basille, B. Van Moorter, Habitat–performance relationships: finding the right metric at a given spatial scale, Phil. Trans. R. Soc. B 365 (1550) (2010) 2255–2265.

[49] Y. Mao, D. Wang, M. Muller, K.R. Varshney, I. Baldini, C. Dugan, A. Mojsilović, How data scientistswork together with domain experts in scientific collaborations: To find the right answer or to ask the right question? Proc. ACM Hum.- Comput. Interact. 3 (GROUP) (2019) 1–23.

[50] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognit. Lett. 31 (8) (2010).

[51] E. Amigó, J. Gonzalo, J. Artiles, F. Verdejo, A comparison of extrinsic clustering evaluation metrics based on formal constraints, Inf. Retr. 12 (2009) 461–486.

[52] S. Gupta, P. Kumar, B. Bhasker, A rough connectedness algorithm for mining communities in complex networks, in: Big Data Analytics and Knowledge Discovery: 18th International Conference, DaWaK 2016, Porto, Portugal, September 6-8, 2016, Proceedings 18, Springer, 2016, pp. 34–48.

[53] S. Gupta, P. Kumar, An overlapping community detection algorithm based on rough clustering of links, Data Knowl. Eng. 125 (2020) 101777.

[54] M. Kukar, I. Kononenko, Reliable classifications with machine learning, in: Machine Learning: ECML 2002: 13th European Conference on Machine Learning Helsinki, Finland, August 19–23, 2002 Proceedings 13, Springer, 2002, pp. 219–231.

[55] P. Vuttipittayamongkol, E. Elyan, A. Petrovski, On the class overlap problem in imbalanced data classification, Knowl.-Based Syst. 212 (2021) 106631.

[56] M. Khushi, K. Shaukat, T.M. Alam, I.A. Hameed, S. Uddin, S. Luo, X. Yang, M.C. Reyes, A comparative performance analysis of data resampling methods on imbalance medical data, IEEE Access 9 (2021) 109960–109975, http://dx.doi.org/10.1109/ACCESS.2021.3102399.

[57] Y. Lucas, J. Jurgovsky, Credit card fraud detection using machine learning: A survey, 2020, arXiv preprint arXiv:2010.06479.

[58] J.M. Johnson, T.M. Khoshgoftaar, Survey on deep learning with class imbalance, J. Big Data 6 (1) (2019).

[59] E.A. Lopez-Rojas, A. Elmir, S. Axelsson, Paysim: A financial mobile money simulator for fraud detection, 2016.

[60] E.A. Lopez-Rojas, S. Axelsson, BankSim: A bank payment simulation for fraud detection research, 2014.

[61] M. Ahmed, M.A. Kashem, M. Rahman, S. Khatun, Review and analysis of risk factor of maternal health in remote area using the internet of things (IoT), in: A.N. Kasruddin Nasir, M.A. Ahmad, M.S. Najib, Y. Abdul Wahab, N.A. Othman, N.M. Abd Ghani, A. Irawan, S. Khatun, R.M.T. Raja Ismail, M.M. Saari, M.R. Daud, A.A. Mohd Faudzi (Eds.), InECCE2019, Springer, Singapore, 2020.

[62] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995).

[63] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inform. Theory 13 (1) (1967) http://dx.doi.org/10.1109/TIT.1967.1053964.

[64] P. Domingos, M. Pazzani, On the optimality of the simple Bayesian classifier under zero-one loss, Mach. Learn. 29 (2) (1997) http://dx.doi.org/10.1023/A:1007413511361.

[65] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016.

[66] P. Schober, C. Boer, L.A. Schwarte, Correlation coefficients: appropriate use and interpretation, Anesth. Analg. 126 (5) (2018) 1763–1768.

[67] A. Chhabra, P. Li, P. Mohapatra, H. Liu, " what data benefits my classifier?" enhancing model performance and interpretability through influence-based data selection, in: The Twelfth International Conference on Learning Representations, 2024.

[68] A. Paullada, I.D. Raji, E.M. Bender, E. Denton, A. Hanna, Data and its (dis) contents: A survey of dataset development and use in machine learning research, Patterns 2 (11) (2021).

[69] R. Decoupes, M. Roche, M. Teisseire, GeoNLPlify: A spatial data augmentation enhancing text classification for crisis monitoring, Intell. Data Anal. (Preprint) (2024) 1–25.

[70] K. Hanauer, M. Henzinger, C. Schulz, Recent advances in fully dynamic graph algorithms–a quick reference guide, ACM J. Exp. Algorithmics 27 (2022) 1–45.

[71] V.S.S. Ram, N. Kayastha, K. Sha, OFES: Optimal feature evaluation and selection for multi-class classification, Data Knowl. Eng. 139 (2022) 102007.

[72] M.R. Bendre, V.R. Thool, Analytics, challenges and applications in big data environment: a survey, J. Manag. Anal. 3 (3) (2016) 206–239.

[73] F.T. Tschang, E. Almirall, Artificial intelligence as augmenting automation: Implications for employment, Acad. Manag. Perspect. 35 (4) (2021) 642–659.

[74] D. Sargiotis, Data security and privacy: protecting sensitive information, in: Data Governance: A Guide, Springer, 2024, pp. 217–245.

[75] M. Abedi, Non-euclidean distance measures in spatial data decision analysis: investigations for mineral potential mapping, Ann. Oper. Res. 303 (1) (2021) 29–50.

[76] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65.

[77] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Comm. Statist. Theory Methods 3 (1) (1974) 1–27.

[78] A.E. Taylor, L'hospital's rule, Am. Math. Mon. 59 (1) (1952) 20–24.

[79] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Information Processing & Management 45 (4) (2009) 427–437.

Payel Sadhukhan is an Associate Professor of Computer Science at Techno Main, Salt Lake, West Kolkata, India. She did her Ph.D. and Masters in Computer Science from Indian Statistical Institute (ISI), Kolkata. Her doctoral work lies in the domain of Machine Learning and Artificial Intelligence. Her present interests include Privacy-preserving Machine Learning, feasible computation under privacy-preserving constraints and explainability and interpretability in ML.

Samrat Gupta is an Associate Professor in the Information Systems area at the Indian Institute of Management Ahmedabad, India. He also serves as a Researcher at the University of Agder, Norway and the University of Economics in Bratislava, Slovakia. He obtained his doctoral fellowship from Indian Institute of Management Lucknow, India. He has extensively published in journals of international repute (including FT-50 and CABS 4/4*) contributing to the areas of network theoretic modelling, information disorder, user engagement on online platforms and user centered digitalization. His research projects have been funded by Ministry of Education, Government of India and European Union's Horizon research and innovation programme.