applied sciences

MDPI

Article

Using Feature Selection with Machine Learning for Generation of Insurance Insights

Ayman Taha [1,2], Bernard Cosgrave [3] and Susan Mckeever [1,*]

[1] School of Computer Science, Technological University Dublin, D02 HW71 Dublin, Ireland; ayman.farahat@tudublin.ie
[2] Faculty of Computers and Artificial Intelligence, Cairo University, Giza 12613, Egypt
[3] DOCOsoft, D03 E5R6 Dublin, Ireland; bernard.cosgrave@docosoft.com
* Correspondence: susan.mckeever@tudublin.ie

Abstract: Insurance is a data-rich sector, hosting large volumes of customer data that is analysed to evaluate risk. Machine learning techniques are increasingly used in the effective management of insurance risk. Insurance datasets by their nature, however, are often of poor quality with noisy subsets of data (or features). Choosing the right features of data is a significant pre-processing step in the creation of machine learning models. The inclusion of irrelevant and redundant features has been demonstrated to affect the performance of learning models. In this article, we propose a framework for improving predictive machine learning techniques in the insurance sector via the selection of relevant features. The experimental results, based on five publicly available real insurance datasets, show the importance of applying feature selection for the removal of noisy features before performing machine learning techniques, to allow the algorithm to focus on influential features. An additional business benefit is the revelation of the most and least important features in the datasets. These insights can prove useful for decision making and strategy development in areas/business problems that are not limited to the direct target of the downstream algorithms. In our experiments, machine learning techniques based on a set of selected features suggested by feature selection algorithms outperformed the full feature set for a set of real insurance datasets. Specifically, 20% and 50% of features in our five datasets had improved downstream clustering and classification performance when compared to whole datasets. This indicates the potential for feature selection in the insurance sector to both improve model performance and to highlight influential features for business insights.

Keywords: feature selection; feature reduction; machine learning; insurance insights; insurance data analytics

Citation: Taha, A.; Cosgrave, B.; Mckeever, S. Using Feature Selection with Machine Learning for Generation of Insurance Insights. *Appl. Sci.* **2022**, *12*, 3209. https://doi.org/10.3390/app12063209

Academic Editor: Andrea Prati

Received: 31 January 2022
Accepted: 11 March 2022
Published: 21 March 2022

1. Introduction

The insurance sector by nature has been an intensively data-driven industry for many years, with insurance companies managing large quantities of customer data. The business of insurance is based on the analysis of data to understand and effectively evaluate risk. The insurer makes use of actuaries and actuarial science techniques to analyse insurance data to perform core roles. Therefore, insurance data can be claimed to be a dominant force in the sector [1]. Enhancing the quality of insurance data through appropriate pre-processing should improve the estimation process by increasing data quality. Examples of pre-processing tasks include handling of missing values [2], managing outliers [3], binning numerical data to create categories [4], and better handling categorical data correlation/association [5].

The use of machine learning in the non-life insurance industry can be divided into three categories: actuarial, fraud detection, and customer behaviour. Actuaries carry out two tasks relevant to machine learning: calculating how much an insurance policy should cost, which is called pricing or *ratemaking*, and calculating how much money an insurer

should set aside for the payment of future claims, which is called *reserving*. The nature of insurance data (especially aggregated data) and the detailed domain requirements means that bespoke domain specific models are required within the actuarial category. Fraud detection and customer behaviour (churn and propensity to buy) are however amenable to the application of more general machine learning techniques, such that advances in other domains can be applied to problems found in these categories. To give an example, techniques for detecting credit card fraud can be used to detect insurance fraud [6] and techniques developed for churning in the supermarket sector can also be applied to the insurance sector [7].

In training datasets for supervised learning, redundant and irrelevant features have been demonstrated to affect the performance of learning models. Taking the commonly used Support Vector Machine (SVM) algorithm, the identification of important features significantly improves the robustness of SVM learning models [8], with such models being sensitive to noisy features. SVM models aim to optimise the kernel of the data, where these models suppose that the kernel matrix of training data is positive definite. However, the positive definite assumption cannot be ensured because of the influence of noisy features. Therefore, it is recommended to decrease the influence of noisy features to build robust SVM-based models [9].

Insurance databases tend to contain multiple redundant and irrelevant attributes. These attributes negatively affect the accuracy of prediction of insurance reserve prediction techniques. Therefore, it is intuitively important to apply/embed feature selection prior to the creation of Machine Learning (ML) models in order to strip out low influence features. Furthermore, improving the model prediction power by feature selection and dimensionality reduction holds promise towards improving the processing and accuracy of many insurance problems such as insurance reserve prediction, customer retention, policy price, and insurance fraud detection. In our related works section we outline both the application of ML in the insurance sector and discuss various techniques used to implement feature selection. Above and beyond improved algorithmic performance, feature selection reveals the most and least influential features in a dataset for the model task in question. Domain experts can use these insights for practical purposes immediately or as a starting point for further investigations [10].

Contributions

This paper highlights the presence and influence of poor data quality on machine learning algorithms using insurance datasets. It brings feature selection and machine learning together in the insurance sector aiming towards the creation of more robust ML models and predictive domain insights. It emphasizes the role of feature selection techniques in improving insurance insights. Our specific contributions are as follows:

- We highlight the influence of noisy, redundant and/or irrelevant features in lessening the accuracy of machine learning algorithms. We show that feature selection can lead to better predictive/classification models in insurance, through removal of noisy features.
- We demonstrate how feature selection can lead to domain insights through examination of the features which most contribute to model decisions.
- We have used the proposed framework to improve the performance of machine learning algorithms on real insurance datasets.

The rest of this paper is organized as follows: Section 2 highlights different applications of machine learning in the insurance sector and reviews feature selection algorithms for mixed data which are suitable for insurance datasets. Our methodology for improving the application of machine learning in the insurance sector is presented in Section 3. Section 4 introduces a comparative study based on benchmark datasets. The discussion is presented in Section 5. Finally, Section 6 concludes the paper.

2. Related Works

This section provides a brief overview of the literature on applications of machine learning in the insurance sector as well as the common feature selection methodologies for mixed datasets.

2.1. Machine Learning Applications in Insurance Sector

As mentioned previously, one of the main applications of machine learning in insurance is in actuarial tasks. While both reserving and pricing actuaries have long made use of linear models, research into the applicability of newer, more advanced, and non-linear techniques has been slow, and it is only relatively recently that this area has attracted much attention [11]. In both ratemaking and reserving, two key motivations behind research into machine learning techniques are (1) the realisation that the factors determining the appropriate rate or reserve are too complex to be modelled in a linear fashion and (2) the availability of more data both in terms of quantity and diversity (including text, image, and telematics) [11–13].

In ratemaking, the traditional method is Generalized Linear Models (GLM) [14,15]. The response variable of these models can have a Gamma or Poisson distribution [16]. This makes them particularly suitable for ratemaking as both claims severity and frequency need to be estimated and claim severity is often Gamma distributed, while the Poisson distribution is used to fit claim frequency [16]. Generalized Additive Models (GAM) go beyond GLMs in the ability to find non-linear relationships and have been explored firstly by Denuit et al. [17] and more recently by Klein et al. [18]. Neural networks have been part of the research into ratemaking for more than 20 years and have also been studied more recently; they show particular promise in modelling non-linear relationships that are not captured by more basic models [11,19,20]. Wong et al. [11] mention an issue we have encountered in our own study; machine learning models, particularly neural networks, require large datasets. However, the confidential nature of insurance data means that large publicly available datasets are few and far between [11].

The other actuarial function, reserving, has been dominated by the chain ladder method. Mack [21] gives a good overview of the early debate on this topic. The key to the chain ladder method is reserve triangles. Reserve triangles are matrices that track accumulating claims data (number of claims and value) over time. A stochastic process is then used to generate the run-off matrices from the reserve triangles. These run-off matrices are used in estimating the final reserve value. Finally, stochastic regression models are used in predicting the insurance reserve values for these claims.

The models aim to predict the future development (frequency and value) of claims based on previous patterns. While the models perform adequately for large claim portfolios, they cannot cope with changes in patterns of behaviour that give rise to claims, whether social, economic, or random. These models also work with aggregated data, which means data related to individual claims or small groups of claims is not taken into account. This in turn means that the output of these models has little use for reserving at an individual claim level [10–12]. In addition, non-structured data which contains information about factors affecting the value of claims and hence the appropriate reserve is lost to the model.

Research has addressed these issues. Machine learning techniques that are not dependent on past behaviour continuing into the future have been applied to reserve prediction. Tree based techniques, SVM, neural networks and deep learning techniques are among the approaches [22–25]. These techniques can also utilise a greater range of features combining numerical, categorical and textual data in the same model. Attention has also been paid to applying machine learning techniques at an individual claim level [26,27].

Insurance fraud detection is another area within the insurance domain in which machine learning can play a part. As mentioned previously, fraud detection in the insurance domain can be viewed as an industry specific application of more general machine learning techniques rather than a unique problem. As discussed by Gomes et al. [6], most industry approaches to this problem understand the problem as a classification problem and

apply supervised learning models in an attempt to classify claims as fraudulent or not. Itri et al. [28] provide good examples of this approach. These approaches face two related problems; first, there is no natural fraud label in insurance claims data, so data must be labelled by hand and this process is subject to error and inconsistency. Second, fraud is a major problem for insurance companies, but the vast majority of insurance claims are not fraudulent, so datasets suffer from a class imbalance. Hassan et al. [29] provide one approach to the problem of class imbalance. An alternative approach is not to aim to improve the existing machine learning approach incrementally but to look to text mining, deep learning, and unsupervised learning. Wang et al. [30] combine text mining with deep learning to propose a framework for the detection of fraudulent motor insurance claims. They first use Latent Dirichlet Allocation on the description of accidents contained in claims document to extract features and then combine these with traditional features in a deep neural network model. They report that their results outperform traditional supervised learning approaches. Gomes et al. [6] utilise a deep neural network in combination with unsupervised techniques. We note, however, the need for very large datasets and the production of opaque black-box models with the use of deep learning.

Another area where there is scope for machine learning is churn prediction or the propensity to buy an insurance product. Again the techniques utilised to tackle this issue in insurance can readily draw from other industries, including financial services. The literature in the field specifically related to insurance, especially if we exclude medical insurance in the USA, is relatively sparse. An overview of how traditional machine learning techniques can be applied in this area is provided in [7,31]. How neural networks can supplement a tree model is shown in [8]. Although they are working with supermarket data, we would expect similar outcomes with insurance data.

2.2. *Feature Selection Techniques*

In this section, we provide an overview of feature selection approaches. We then give a more detailed description for the specific feature selection methods used in our study.

Feature selection techniques based on the selection strategy can be classified into three types: wrapper, filter and embedded [32].

Wrapper methods employ a predefined supervised or unsupervised learning algorithm to find the best subsets of features that improve the classification accuracy or clustering quality of corresponding learning algorithms [33,34]. Therefore, they are dependent on the learning algorithm. Wrapper feature selection techniques often get high performance measures but they consume high computational time to run. Moreover, the group of selected features for a certain dataset may be changed if another learning algorithm is used. They are limited to being employed in conjunction with a specified learning algorithm [35]. Filter feature selection techniques are independent of the machine learning algorithm used. They use general characteristics of features such as variance, consistency, correlation, and information to pick features without involving any machine learning algorithms [36]. Despite the fact that filter based feature selection methods do not often outperform wrapper-based methods, they are widely used in the literature because they are typically fast, scalable, and applicable to high-dimensional data. Furthermore, they are usually performed before classification, and thus they they are not biased towards learning algorithms [4]. Embedded techniques combine characteristics of both filter and wrapper methods. They usually employ feature selection in conjunction with predefined learning process similar to wrapper methods but they do not iterate the learning algorithm, which is in contrast to original wrapper methods. Therefore, they are more efficient than wrapper methods but they do not typically outperform them [37,38].

In this article, we choose filter based feature selection methods to demonstrate and investigate the potential for feature selection on insurance datasets, due to their wide use in the literature and independence of the learning algorithm used.

Structured insurance datasets are typically mixed containing a mix of numeric and categorical data so techniques suited to mixed data will be considered.

Filter based feature selection methods make use of some intrinsic properties to compute a score for each feature. These methods evaluate the relevance of the features either individually or jointly. They usually require the identification of the target number of selected features or the desired percentage of total number of features such as 0.2, 0.5, and 0.8. There are two methodologies for selecting features: ranking based methods or selection based methods. The selection based features selection is an iterative method wherein the algorithm starts with the best performing feature in the result set. Next, it adds another feature that gives the best performance in combination with the features in the result set until the desired number of features is reached. However, ranking based methods compute a relevance score for each feature, ranking all features based on this score. Next, they choose the top percentage of features that have the best scores. To demonstrate the potential for feature selection, we need to use several methods to remove reliance on any single set of results. We select the following four feature selection techniques that are suitable for mixed datasets: the Greedy Feature Selection Algorithm (GFSA) [39], the Laplacian Score (LS) [40], the spectral algorithm (Spec) [41], and the Unsupervised Spectral Feature Selection Method (USFSM) [42]. We will explain each of these below.

Before moving on to explain the four techniques that we have used, we highlight some new approaches that stand out as interesting. Since our purpose in this paper is to demonstrate the value of feature selection for insurance domain and not to assess state of art feature selection techniques, we will only briefly outline these developments. The first Paniri et al. paper [43] developed a filter feature selection algorithm based on Ant Colony Optimisation (ACO). Their approach specially designed for multi-labelled data uses artificial agents called ants whose behaviour is modelled on real ants. As the ants traverse the feature space they leave a pheromone trail which is updated in each iteration. At the end of all iterations the top k features with the highest pheromone intensity are selected. The search for the most informative features is modelled as a graph with nodes representing features. In contrast to previous ACO approaches where the initial pheromone value for each node is a constant, the authors here initialise the pheromone values by calculating the normalized cosine similarity between each feature and class label. Ants are randomly placed on nodes and the next step of the ants is decided by a decision policy. Interestingly, the heuristics involved in the decision making take account of both redundancy, the correlation between the features, and relevancy, the correlation between features and class labels. The pheromone values of each node are then updated and the process is repeated until the maximum number of iterations is reached.

Recent ensemble-based feature selection methods are proposed: Ensemble of Feature Selection Multi-Criteria Decision-Making (EFS-MCMD) algorithm [44] and Pareto-based Ensemble of Feature Selection (PEFS) algorithm [45]

In EFS-MCMD, the authors combine the understanding of how it may be possible to improve the robustness of feature selection methods through using more than one algorithm with the idea that using the Multi-Criteria Decision-Making (MCDM) process may get the best results from the ensemble model. Ensemble methods aim to overcome the problem of a single feature selection algorithm finding only a local optima of selected features from the whole feature space. Combining the output of multiple feature selection algorithms can overcome this problem and improve accuracy. Ensemble feature selection methods can be divided into two categories: homogeneous and heterogeneous. Homogeneous approaches partition the data into subsets and apply a single feature selection algorithm to each subset and then combine the ranking from each subset. Heterogeneous approaches apply multiple feature selection algorithms to the whole dataset and combine the results of each algorithm into a final ranking. The authors use a heterogeneous approach.

EFS-MCDMs are methods where multiple and sometimes conflicting opinions and criteria are weighed to reach the best alternatives and are particularly useful for solving problems where ranking is required. The authors use the VIKOR (from Serbian: Vlsekriterijumska Optimizacija I Kompromisno Resenje) method for solving the MCDM problem which involves constructing a decision matrix. This implementation of the VIKOR method

takes the proposed top k most relevant features of each feature selection algorithm as criteria for the decision matrix. The VIKOR method ranks the features by the VIKOR index value. The method proceeds in the following way. The user selects the number of features required, the ensemble feature selection algorithm is run, and then the VIKOR method is applied to find the most relevant features. The results based on experiments with real-world datasets are impressive in terms of accuracy, f-score, and run-time.

Another ensemble feature selection algorithm is PEFS [45]. It employs multiple filter based feature selection algorithms and multi-objective optimization techniques to obtain the final optimal feature subset (dominant solution). If there is no single optimal solution, the feature selection algorithms in PEFS deliver a set of trade-off solutions called the Pareto optimal set. These solutions are called non-dominated solutions. PEFS formulates the problem of feature selection as a Pareto-based optimization problem with two objectives. It employs four different feature selection algorithms and then combines their results using two aggregation methods. It evaluates these results using a bi-objective optimization. The features are ranked according to their crowding distance in the bi-objective space. In multi-objectives optimization literature, the crowding distance of a solution estimates the density of surrounding solutions [46]. PEFS utilises a combination of redundancy and relevancy feature selection techniques to give a higher ranking to the features which have the higher relevancy score and the lowest redundancy score.

Here, we discuss the candidate feature selection algorithms for use in our work. LS and Spec algorithms are feature selection techniques for continuous datasets. Thus, each nominal feature is replaced by an ordinal feature. Each category is transformed by an integer that represents its order of appearance within this column. We have chosen these methods because they are widely used in the literature because they are typically scalable and applicable to high-dimensional data [42].

The Greedy Feature Selection Algorithm (GFSA) [39] is based on selecting the set of features that has the minimum association amongst them. This algorithm searches for a diverse and minimally dependent set of features. It formulates the feature selection problem as an optimization problem. GFSA handles categorical as well as numerical features. The GFSA algorithm follows a step-by-step procedure to find the desired dimension that minimizes the multiple associations among them. It starts with choosing the least dependent pair of features. Then it increases the solution by one feature at each iteration until it reaches the desired number of features. The GFSA algorithm assumes that the incoming feature should have the minimum pair-wise association with one feature in the solution set. It then computes the multiple association for all candidate sets and selects the set with the least multiple association [39].

The Laplacian Score (LS) [40] is a filter based feature selection algorithm. The LS feature selection is designed only for continuous data. The Eigensystem of the Laplacian matrix is employed to measure the relevance of each feature. The Laplacian matrix is computed from the similarities among objects. The features are ranked according to their Laplacian scores which reflect locality preserving power of a single feature. The LS score relies on the geometric structure of objects. At the beginning, the LS algorithm builds a nearest neighbour graph G. Each node represents a feature. Two nodes are IFF and one node is among the k nearest neighbours of the other node. It then computes a Laplacian score for each feature based on the k nearest neighbour graph, which is used a measure its power of locality preserving. The LS looks for those features that respect this graph structure [40].

The spectral feature selection method, Spec, [41] relies on the spectral graph theory. The Spec algorithm utilizes the intrinsic properties to construct a similarity graph, G, to represent the dataset. The relevance of features is measured by their consistency with the structure of the similarity graph among objects. Moreover, a similarity score for each feature is computed. This score should be consistent with the graph structure. The Spec algorithm computes a consistency score for each feature to measure the coherence between this feature and the nontrivial eigenvectors of the Laplacian matrix.

The Unsupervised Spectral Feature Selection Method (USFSM), is an unsupervised feature selection technique for mixed data. It is based on building a similarity matrix between objects using a kernel function. The kernel function makes use of distance functions for numeric and nominal features. The structural information of the data is contained in the spectrum of the Laplacian matrix through the separation of its elements.

This algorithm determines the relevance of a feature by measuring the change of spectrum distribution of the Normalized Laplacian matrix when a feature is excluded. A leave-one-out strategy is iteratively used to compute a spectral gap score for each feature. It then ranks features based on spectral gap score.

In the next section, we explain our framework for applying feature selection.

3. Proposed Framework

Our aim is to demonstrate and investigate feature selection on insurance datasets. We present our approach to applying feature selection as a framework, explained in two parts: defining the problem statement and presenting the proposed framework.

3.1. Problem Statement

Finding the most influential insurance features can be formally expressed as follows:

Given:

- A insurance dataset, D, consisting of $v = p + q$ mixed features, where
  - p: The number of quantitative features and;
  - q: The number of qualitative features.

Find:

- The best representative set of features.

Objective

- Identify irrelevant and redundant feature.
- Improve machine learning accuracy.

3.2. Framework Description

Our proposed framework aims to highlight the importance of employing feature selection algorithms prior to applying machine learning algorithms for insurance data. The steps of the proposed framework are shown in Algorithm 1.

Input:

- D: A Mixed Insurance Dataset containing v Mixed features.
- C: List of candidate feature selection algorithms.
- π: predefined set of features percentages.

Output:

- S: The best representative set of features to represent D.

Algorithm 1 Proposed framework for identifying most influential attributes in insurance datasets

```
for < each feature selection algorithm c_i in C> do
    for < each each percentage π_j of selected feature in π> do
        D_ij ← Identify selected feature by Algorithm c_i
        Apply intended machine learning algorithm on D_ij
        Compute performing measures
    end for
end for
Identify D_ij that gets best performance results
return D_ij
```

The proposed framework requires the input datasets, the list of numbers of desired features, and the choice of candidate feature selection algorithms. For each pair of feature selection algorithm and percentage of feature selection, the set of classification and clustering-based performance measures is computed. It then returns the set of features of required size that has the best performance measures.

4. Methods and Materials

In this section, we aim to discover insurance insights and improve the performance of machine learning in the insurance sector through examining the effect of our proposed framework to highlight the importance of feature engineering in focusing, ranking, and selecting influential features. We examine the candidate feature selection methods based on the accuracy measures of machine learning algorithms before and after applying the feature selection algorithm on insurance datasets to highlight the importance of the usage of feature engineering, before applying machine learning techniques in the insurance sector.

In these experiments, parameter settings for comparative published methods were fixed according to the recommendation of their respective authors. We implement GFSA in the R language. Fernandez et al. [42] kindly provided us with their Java program for computing USFSM. Both the LS and Spec algorithms are provided in an R-package named "Rdimtools" for Dimension Reduction and Estimation Methods. All experiments were run in R, using a computer with an Intel Core i7-6600U 2.60 GHz processor with 16 GB DDR4 RAM, running 64-bit Windows 10.

4.1. Experimental Design

We propose an experimental design for highlighting the effect of employing feature selection algorithms before applying machine learning tasks for insurance. All experiments were run on a computer with an Intel Core i7-6600U processor running at 2.60 GHz using 16 GB of RAM, running Windows 10. In these experiments, we make use of the following candidate feature selection algorithms:

1. GFSA: The Greedy Feature Selection Algorithm [39];
2. LS: The Laplacian Score [40];
3. Spec: The Spectral Algorithm [41] ;
4. USFSM: The Unsupervised Spectral Feature Selection Method [42].

For each candidate algorithm, we have applied different proportions of selected features, $\pi = \{0.2, 0.5, \text{and } 0.8\}$. All algorithms apart from GFSA are ranked-based feature selection algorithms, which means that they rank all features according to their relevance. We then use this ranking to identify the top π percentage of these features. GFSA, however, requires the number of desired features, M, in advance as an input parameter. Therefore, we run GFSA with each value of π that we wish to investigate. We have 12 configurations, consisting of 4 algorithms across 3 proportions of selected features. We compute the downstream evaluation measures for each configuration. These evaluation measures are also computed for whole datasets (all features) to provide our baseline without the use of FS selection. Finally, we select the set of features that has the best performance measures.

We would like to mention that all candidates solutions, except GFSA, are sensitive to the number of objects (observations) in the dataset. These techniques are challenging to run when the number of observations is high because their time complexity increases non-linearly as the number of observations increases. These algorithms, except GFSA, could not be applied for long datasets such as D4 in Table 1.

Table 1. Insurance datasets' description.

Id	Dataset Name	No. of Columns			No. of Rows	No. of Classes
		Numerical	Categorical	Total		
D1	Car Insurance Cold Calls	10	8	18	4000	2
D2	Medical insurance claim fraud	5	7	12	7000	2
D3	Caravan Insurance Challenge	76	10	86	9822	2
D4	Health_Insurance_Lead_Prediction	6	6	12	50,882	2
D5	Insurance company-Customers willing to buy a new product	10	3	13	14,017	2

4.2. Real Insurance Datasets

For our work, we looked for insurance benchmark datasets in the public machine learning repositories, e.g., UCI [47] and Kaggle [48]. We excluded many insurance datasets because they are either not annotated or lacked meta information (such as feature titles) for interpretation, finally selecting five insurance datasets for this comparison. These datasets are publicly available from the Kaggle machine learning repository [48]. Furthermore, we manually prune the features to remove individual identifiers or constant features. Table 1 shows the characteristics of selected datasets, describing each of the datasets as set out in the repository [48].

The first dataset, Car Insurance Cold Calls, is a dataset from a bank, which constructs campaigns to engage new clients for car insurance services. The bank needs to predict whether those potential customers will buy car insurance or not based on their data from previous campaigns. It provides general information about clients and specific information about the insurance campaigns. This dataset consists of 18 attributes (after removing the id attribute) in relation to 4000 customers contacted during the last campaign.

The Medical Insurance Claim Fraud dataset relates to the detection of health insurance fraud claims based on patients' data. It provides general information about insurance claims and their owners such as gender, location, employer, cause, and fee charged. The original datasets contain 15 attributes and 7000 claims. However, we removed 3 noisy attributes: patient name, DOB, and email.

The Caravan Insurance Challenge dataset was used in the CoIL 2000 challenge [49]. It contains socio-demographic information about clients of an insurance company to predict which potential customer will buy a caravan insurance policy. It consists of 86 attributes and 9822 observations.

The Health Insurance Lead Prediction dataset was collected by a financial services company. This company built this dataset to predict whether a client is interested in its recommended medical insurance policies based on their profile on the company website. The client information includes demographics information and previous holding polices. The customer is classified as a lead if she/he fills in the policy form. Its training dataset consists of 50,882 rows and 13 attributes after removing the individual identifier attribute for prediction, "Id".

The final dataset is named Insurance Company. It aims to identify customers willing to buy a new insurance product. It contains customer information and their buying behaviour of two other services. It consists of 15 attributes and 14,017 customers. We removed two features, customer-id and contract, because contract contains a constant value which has 0 variance, so it cannot be used for extracting any interesting information.

4.3. Evaluating Measures

We have chosen two types of machine learning tasks: classification-based approach for supervised learning and clustering-based approach for unsupervised learning, to demonstrate the effect of features selection algorithm in machine learning for insurance datasets. We followed the standard methodology for evaluating feature selection methods [42]. To evaluate the clustering performance of insurance datasets before and after feature selection, we utilise the well-known clustering algorithm for clustering mixed data k-

prototypes [50]. In the comparison, two clustering measures are used: clustering accuracy (C-ACC) and normalized mutual information (NMI). The C-ACC is calculated as follows

$$\text{C-ACC}(\theta) = \sum -i = 1^{n} \frac{\delta(p, map(q_i))}{n}, \quad (1)$$

where n is the total number of observations and $\delta(a, b) = 1$ if $a = b$: otherwise $\delta(a, b) = 0$, and $map(q_i)$ is a mapping function that permutes clustering labels to get the best match with the true labels based on the Kuhn–Munkres algorithm [51]. The C-ACC values range from 0 to 1; the clustering is better when the C-ACC is higher.

For clustering results (P) and true labels (Q), NMI [40] is defined as:

$$\text{NMI}(\theta) = \frac{I(P, Q)}{max(H(P), H(Q))}, \quad (2)$$

where $H(P)$ and $H(Q)$ are the entropies of P and Q, respectively, and $I(P, Q)$ is the mutual information [52] between P and Q, which is defined as:

$$I(P, Q) = \sum_{p_i \in P \wedge q_j \in Q} p(p_i, q_j)\ log_2 \frac{p(p_i, q_j)}{p(p_i)\ p(q_j)}. \quad (3)$$

In the classification evaluation-based approach, two classification techniques, Support Vector Machine (SVM) and K-Nearest Neighbours (KNN), are used to evaluate classification accuracy: the ratio between the correctly classified objects and the total number of objects.

4.4. Experimental Results Analysis

The clustering accuracy for selected datasets is shown in Table 2. For each dataset, D_i in the table displays 3 groups (rows) of results. Each row indicates results corresponding to a percentage of selected features, $\pi = \frac{M}{v}$, is set to 0.2, 0.5, and 0.8. As shown in Table 2, there is at least one subset of features that has better clustering accuracy than the whole dataset. This indicates the existence of noisy features, which reduces the clustering accuracy of the whole dataset. If we remove these noisy features the performance will improve. For example, with D1, 9 out of 12 configurations using FS algorithms lead to higher clustering accuracy than using the full dataset. The best clustering accuracy for D1, 0.6613, is found when we select 0.5 of features by Spec Algorithm. The best clustering accuracy for D2 and D3 occurs when we select 0.2 of features by Spec Algorithm. We found that if we select 0.2 of features in D1, D2, D3, and D5, by Spec Algorithm, we will get better clustering accuracy than by applying the clustering algorithm to the whole dataset.

Table 3 shows the NMI results for the selected 5 datasets. Similar to clustering accuracy, for all datasets, there is at least one subset of features which outperforms using the whole dataset in terms of NMI. This emphasizes the existence of noisy features that negatively affect the NMI accuracy of the whole dataset. For example, all subsets of features, recommended by all algorithms in D1, have NMI accuracy better than the whole dataset. The best NMI accuracies are found when we select 0.5 percentage of features by Spec Algorithm. Corresponding with clustering accuracy results, we find that if we select 0.2 of features in D1, D2, D3, and D5 by Spec Algorithm, we will get better NMI accuracy than by applying the clustering algorithm to whole datasets.

The Spec algorithm has the best clustering-based performing measures for our five insurance datasets.

Table 2. Clustering accuracy where four feature selection (FS) methods were applied, showing results for three feature percentages versus all features. Highest FS-based result is in bold.

Dataset	Per. of. Feat.	No. Sel. Col	GFSA	LS	Specu	USFSM	All-Features
D1	0.2	4	0.4018	0.6600	0.6600	0.5373	0.5713
	0.5	8	0.5983	0.5600	**0.6613**	0.5528	
	0.8	14	0.5720	0.5713	0.6378	0.5898	
D2	0.2	18	0.3188	0.5370	**0.5408**	0.3117	0.4045
	0.5	44	0.3145	0.3151	0.3264	0.3152	
	0.8	70	0.3143	0.4045	0.3157	0.3153	
D3	0.2	3	0.3201	0.3207	**0.4512**	0.3207	0.3198
	0.5	7	0.3198	0.3426	0.3168	0.3746	
	0.8	11	0.3198	0.3198	0.3148	0.3198	
D4	0.2	3	0.5307	Mem Ex	Mem Ex	Mem Ex	0.3369
	0.5	7	**0.5357**	Mem Ex	Mem Ex	Mem Ex	
	0.8	11	0.3376	Mem Ex	Mem Ex	Mem Ex	
D5	0.2	2	0.4989	0.4930	0.5351	0.5220	0.4964
	0.5	5	0.4989	0.5389	0.4959	0.4940	
	0.8	8	0.5000	**0.6307**	0.4964	0.4964	

Table 3. NMI accuracy for insurance datasets.

Dataset	Per. of. Feat.	No. Sel. Col	GFSA	LS	Specu	USFSM	All-Features
D1	0.2	4	0.0301	0.0528	0.0528	0.0077	0.0090
	0.5	8	0.0301	0.0045	**0.0541**	0.0056	
	0.8	14	0.0082	0.0090	0.0418	0.0261	
D2	0.2	18	0.0059	0.0033	**0.0030**	0.0064	0.0038
	0.5	44	0.0070	0.0065	0.0066	0.0065	
	0.8	70	0.0073	0.0038	0.0066	0.0065	
D3	0.2	3	0.0865	0.0848	**0.1001**	0.0848	0.0871
	0.5	7	0.0871	0.0669	0.0903	0.0467	
	0.8	11	0.0871	0.0871	0.0927	0.0871	
D4	0.2	3	0.0000	Mem Ex	Mem Ex	Mem Ex	0.0001
	0.5	7	0.0000	Mem Ex	Mem Ex	Mem Ex	
	0.8	11	**0.0001**	Mem Ex	Mem Ex	Mem Ex	
D5	0.2	2	0.0001	0.0000	0.0001	0.0000	0.0000
	0.5	5	0.0001	0.0038	0.0000	0.0000	
	0.8	8	0.0000	**0.0294**	0.0000	0.0000	

Table 4 shows the KNN classification accuracy. Like clustering accuracy and KNN accuracy, for all datasets, there is at least one subset of features which outperforms the whole dataset in terms of KNN accuracy. This emphasizes the existence of noisy features that negatively affect the KNN accuracy of whole datasets.

Table 4. KNN Accuracy for insurance datasets.

Dataset	Per. Of. Feat.	No. Sel. Col	GFSA	LS	Specu	USFSM	All-Features
D1	0.2	4	0.6648	0.6295	0.6325	0.6613	0.6588
	0.5	8	0.6430	0.6530	0.6715	**0.6760**	
	0.8	14	0.6458	0.6695	0.6563	0.6710	
D2	0.2	18	0.9378	**0.9396**	0.9388	0.9385	0.9347
	0.5	44	0.9352	0.9385	0.9385	0.9380	
	0.8	70	0.9343	0.9345	0.9334	0.9384	
D3	0.2	3	0.6814	0.6828	0.6805	0.6831	0.9166
	0.5	7	0.8931	0.6934	0.6938	0.7394	
	0.8	11	0.8894	0.9135	**0.9252**	0.8950	
D4	0.2	3	0.7336	Mem Ex	Mem Ex	Mem Ex	0.7319
	0.5	7	0.7325	Mem Ex	Mem Ex	Mem Ex	
	0.8	11	**0.7338**	Mem Ex	Mem Ex	Mem Ex	
D5	0.2	2	**0.8030**	**0.8030**	0.7794	**0.8030**	0.7814
	0.5	5	0.7779	0.7767	0.7789	0.7789	
	0.8	8	0.7819	0.7817	0.7794	0.7799	

Table 5 depicts the SVM classification accuracy. Like clustering accuracy and SVM classification accuracy, for all datasets except D3, there is at least one subset of features which outperforms using the whole dataset. This emphasizes the existence of noisy features that negatively affect the KNN accuracy of the whole dataset.

The only exception case in our experiments is the SVM classification accuracy for D3. We found that the best SVM accuracy occurred for the whole dataset. There is no subset that outperforms SVM classification accuracy using all features in D3.

Table 5. SVM accuracy for insurance datasets.

Dataset	Per. of. Feat.	No. Sel. Col	GFSA	LS	Specu	USFSM	All-Features
D1	0.2	4	0.6035	0.5890	0.5893	0.6020	0.6633
	0.5	8	0.6448	0.6635	0.6598	0.6610	
	0.8	14	**0.6710**	0.6640	0.6618	0.6660	
D2	0.2	18	0.9403	0.9403	0.9403	0.9403	0.9403
	0.5	44	0.9403	0.9403	0.9403	0.9403	
	0.8	70	0.9403	0.9403	0.9403	0.9403	
D3	0.2	3	0.6799	0.5708	0.5708	0.5708	**0.6829**
	0.5	7	0.6792	0.6799	0.6800	0.6799	
	0.8	11	0.6787	0.6823	0.6792	0.6810	
D4	0.2	3	0.7601	Mem Ex	Mem Ex	Mem Ex	0.7601
	0.5	7	0.7601	Mem Ex	Mem Ex	Mem Ex	
	0.8	11	**0.7601**	Mem Ex	Mem Ex	Mem Ex	
D5	0.2	2	0.8030	0.8030	0.8030	0.8030	0.8030
	0.5	5	0.8030	0.8030	0.8030	0.8030	
	0.8	8	0.8030	0.8030	0.8030	0.8030	

5. Discussion

The objective of this article is to highlight the benefit of feature selection for insurance insights and learning-based tasks. We have proposed a framework for selecting the most influential and discarding irrelevant and noisy features.

We applied our framework to identify the most and least influential feature when applying predictive or descriptive analytical algorithms. In order to demonstrate our claim, we have collected and analysed the results of real insurance datasets publicly available in Kaggle [48]. In most cases (all cases except one), there is at least one subset of features which outperforms all features. This indicates that our benchmark insurance datasets contain irrelevant features that degrade the performance of classification and clustering algorithms.

5.1. Feature Selection Methodologies

Feature selection reduces the number of given features before developing machine learning models. It aims to remove irrelevant, redundant, and/or noisy features to both lessen the processing time of modelling and improve the accuracy of the models. Feature selection techniques can be classified into three types: filter, wrapper, and embedded. We opted to use filter based feature selection methods because they are typically fast, scalable, and applicable to high-dimensional data. In addition, they are widely used in the literature and independent of learning algorithms, such that their results do not changed according to the learning algorithm. In most of the feature selection algorithms, all candidates solutions, except GFSA, are time sensitive to the number of objects (observations). These techniques suffer from computational time issues when the number of observations is high because their time complexity increases non-linearly as number of observations increases. As a result, GFSA is the only algorithm which we could apply to the Health Lead Prediction (D4) dataset. Generally, if we compare four algorithms versus all features, we find that the set of features suggested by Spec algorithm was the best in the majority of cases, especially for clustering-based evaluation measures like clustering and NMI accuracy.

5.2. Generating Insurance Insights

We mentioned previously that a potential benefit of feature selection for insurance datasets is the revelation of the most important and least important features, as a pre-step to generating business insights. Our experimental results from use of our framework revealed the extent to which redundant and irrelevant features exist in the datasets. By identifying which features are influencing (or not) the downstream accuracy gains, we can use these feature level insights to assist business decision making.

From our experimental results, two examples stand out in particular. Insurance companies tend to have more than one line of business to sell across these lines. In the car Insurance dataset, D1, the least important feature is HHInsurance, which indicates whether the target of the call has house insurance with this insurance provider. This immediately indicates that targeting existing customers with house insurance will be of little reward. Balance (yearly average bank balance) and PrevAttempts (no. of previous contacts prior to this campaign) are the second and third least important features respectively. These findings indicate that targeting customers with a high bank balance will be of little reward and that customers do not necessarily tire of being contacted multiple times. The most important features in the D1 dataset are CallStart (start time of the last call), CallEnd (end time of the last call), and Communication (contact communication type). This knowledge enables the company to target particular times of the day to call, choose the particular communication type that works best, and tailor call scripts to favour long or short calls as appropriate.

In the Medical Insurance Claim Fraud dataset, D2, the most important features are gender, location, and number of claims. Knowledge of these factors will allow investigators to target groups of claims for further examination. Fee charge and cause are the least important features and this indicates that the value of the claim and the cause giving rise to medical treatment are not relevant and that focusing on a particular type of claim or value of claim is not likely to uncover fraud.

6. Conclusion and Future Work

The insurance sector has become an intensively data-driven industry, enabling it to offer services that are more responsive to customer needs. Accurate information and insights are required to support decision making. Insurance datasets usually contain irrelevant, noisy, and/or redundant features. These features can negatively affect machine learning techniques, suggesting their removal prior to applying any data analytics algorithms. We propose a framework for selecting the most influential features before applying predictive or descriptive analytical algorithms. We also demonstrate how the feature selection process itself, apart from its role in improving downstream algorithmic performance, can provide insights for insurers that can lead to practical action. The experiments based on real insurance datasets indicate that the application of machine learning techniques based on a set of selected features suggested by feature selection algorithms outperforms the application without employing feature selection. In some cases, we found that half of the features or more in insurance datasets are redundant and irrelevant.

Author Contributions: Conceptualization, A.T. and S.M.; methodology, A.T. and B.C.; formal analysis, A.T. and B.C.; investigation, A.T. and B.C.; writing—original draft preparation, A.T., B.C. and S.M.; writing—review and editing, B.C. and S.M.; supervision, S.M.; project administration, S.M.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This publication has emanated from research supported in part by a grant from Science Foundation Ireland under grant number 18/CRT/6222.

Acknowledgments: Ayman Taha is funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Co-funding of regional, national and international programmes (grant agreement No. 713654).

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

References

1. Hussain, K.; Prieto, E. Big data in the finance and insurance sectors. In *New Horizons for a Data-Driven Economy*; Springer: Cham, Switzerland, 2016; pp. 209–223.
2. Johnson, T.F.; Isaac, N.J.; Paviolo, A.; González-Suárez, M. Handling missing values in trait data. *Glob. Ecol. Biogeogr.* **2021**, *30*, 51–62. [CrossRef]
3. Taha, A.; Hadi, A.S. A general approach for automating outliers identification in categorical data. In Proceedings of the ACS International Conference on Computer Systems and Applications (AICCSA), Ifrane, Morocco, 27–30 May 2013; pp. 1–8.
4. Tang, C.; Liu, X.; Li, M.; Wang, P.; Chen, J.; Wang, L.; Li, W. Robust unsupervised feature selection via dual self-representation and manifold regularization. *Knowl. Based Syst.* **2018**, *145*, 109–120. [CrossRef]
5. Taha, A.; Hadi, A.S. Pair-wise association measures for categorical and mixed data. *Inf. Sci.* **2016**, *346*, 73–89. [CrossRef]
6. Gomes, C.; Jin, Z.; Yang, H. Insurance fraud detection with unsupervised deep learning. *J. Risk Insur.* **2021**, *88*, 591–624. [CrossRef]
7. Scriney, M.; Nie, D.; Roantree, M. Predicting customer churn for insurance data. In *International Conference on Big Data Analytics and Knowledge Discovery*; Springer: Cham, Switzerland, 2020; pp. 256–265.
8. Hu, R.; Zhu, X.; Zhu, Y.; Gan, J. Robust SVM with adaptive graph learning. *World Wide Web* **2020**, *23*, 1945–1968. [CrossRef]
9. Hu, R.; Zhang, L.; Wei, J. Adaptive Laplacian Support Vector Machine for Semi-supervised Learning. *Comput. J.* **2021**, *64*, 1005–1015. [CrossRef]
10. Taha, A.; Cosgrave, B.; Rashwan, W.; Mckeever, S. Insurance Reserve Prediction: Opportunities and Challenges. In Proceedings of the International Conference on Computational Science & Computational Intelligence, Krakow, Poland, 16–18 June 2021; pp. 1–6.
11. Blier-Wong, C.; Cossette, H.L.M.E. Machine Learning in P&C Insurance: A Review for Pricing and Reserving. *Risks* **2020**, *9*, 4.
12. Avanzi, B.; Taylor, G.; Vu, P.A.; Wong, B. Stochastic loss reservingwith dependence: A flexible multivariate tweedie approach. *Insur. Math. Econ.* **2016**, *71*, 63–78. [CrossRef]
13. Dugas, C.; Bengio, Y.; Chapados, N.; Vincent, P.; Denoncourt, G.; Fournier, C. Statistical Learning Algorithms Applied to Automobile Insurance Ratemaking. *Casualty Actuar. Soc. Forum* **2003**, *1*, 179–213.
14. Haberman, S.; Renshaw, A.E. Genaralized linear models and actuarial science. *Statistician* **1996**, *45*, 407–436. [CrossRef]
15. *Generalized Linear Models for Insurance Data*; Cambridge University Press: Cambridge, UK, 2008.
16. Staudt, Y.; Wagner, J. *Comparison of Machine Learning and Traditional Severity-Frequency Regression Models for Car Insurance Pricing*; Technical Report, Working Paper; University of Lausanne: Lausanne, Switzerland, 2019.
17. Denuit, M.; Lang, S. Non-life rate-making with Bayesian GAMs. *Insur. Math. Econ.* **2004**, *35*, 627–647. [CrossRef]
18. Klein, N.; Denuit, M.; Lang, S.; Kneib, T. Nonlife ratemaking and risk management with Bayesian generalized additive models for location, scale, and shape. *Insur. Math. Econ.* **2014**, *55*, 225–249. [CrossRef]

19. Wuthrich, M.V. From Generalized Linear Models to Neural Networks, and Back. Available at SSRN 3491790. 2019. Available online: https://owars.info/mario/2020_Wuthrich.pdf (accessed on 15 January 2022).
20. Wüthrich, M.V.; Merz, M. Yes, we CANN! *ASTIN Bull. J. IAA* **2019**, *49*, 1–3. [CrossRef]
21. Mack, T. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bull. J. IAA* **1993**, *23*, 213–225. [CrossRef]
22. Lopez, O.; Milhaud, X.; Thérond, P.E. Tree-based censored regression with applications in insurance. *Electron. J. Stat.* **2016**, *10*, 2685–2716. [CrossRef]
23. Kuo, K. DeepTriangle: A deep learning approach to loss reserving. *Risks* **2019**, *7*, 97. [CrossRef]
24. Wüthrich, M.V. Neural networks applied to chain–ladder reserving. *Eur. Actuar. J.* **2018**, *8*, 407–436. [CrossRef]
25. Lopes, H.; Barcellos, J.; Kubrusly, J.; Fernandes, C. A non-parametric method for incurred but not reported claim reserve estimation. *Int. J. Uncertain. Quantif.* **2012**, *2*, 39–51. [CrossRef]
26. Wüthrich, M.V. Machine learning in individual claims reserving. *Scand. Actuar. J.* **2018**, *2018*, 465–480. [CrossRef]
27. Kuo, K. Individual claims forecasting with Bayesian mixture density networks. *arXiv* **2020**, arXiv:2003.02453.
28. Itri, B.; Mohamed, Y.; Mohammed, Q.; Omar, B. Performance comparative study of machine learning algorithms for automobile insurance fraud detection. In Proceedings of the 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, 28–30 October 2019; pp. 1–4.
29. Hassan, A.K.I.; Abraham, A. Modeling insurance fraud detection using imbalanced data classification. In *Advances in Nature and Biologically Inspired Computing*; Springer: Cham, Switzerland, 2016; pp. 117–127.
30. Wang, Y.; Xu, W. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decis. Support Syst.* **2018**, *105*, 87–95. [CrossRef]
31. Günther, C.C.; Tvete, I.F.; Aas, K.; Sandnes, G.I.; Borgan, Ø. Modelling and predicting customer churn from an insurance company. *Scand. Actuar. J.* **2014**, *2014*, 58–71. [CrossRef]
32. Solorio-Fernández, S.; Carrasco-Ochoa, J.A.; Martínez-Trinidad, J.F. A review of unsupervised feature selection methods. *Artif. Intell. Rev.* **2020**, *53*, 907–948. [CrossRef]
33. Arai, H.; Maung, C.; Xu, K.; Schweitzer, H. Unsupervised feature selection by heuristic search with provable bounds on suboptimality. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-17), Phoenix, AZ, USA, 12–17 February 2016; pp. 666–672.
34. Guo, J.; Zhu, W. Dependence guided unsupervised feature selection. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-17), New Orleans, LA, USA, 2–7 February 2018; pp. 2232–2239.
35. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection: A Data Perspective. *ACM Comput. Surv.* **2017**, *50*, 94:1–94:45. [CrossRef]
36. Farahat, A.K.; Ghodsi, A.; Kamel, M.S. An efficient greedy method for unsupervised feature selection. In Proceedings of the IEEE International Conference on Data Mining (ICDM), Vancouver, BC, Canada, 11–14 December 2011; pp. 161–170.
37. Wang, S.; Tang, J.; Liu, H. Embedded Unsupervised Feature Selection. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 1–7.
38. Ang, J.C.; Mirzal, A.; Haron, H.; Hamed, H.N.A. Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, *13*, 971–989. [CrossRef]
39. Taha, A.; Hadi, A.S.; Cosgrave, B.; Mckeever, S. A Multiple Association-Based Unsupervised Feature Selection Algorithm for Mixed Data Sets. *Expert Syst. Appl.* **2022**, 1–31.
40. He, X.; Cai, D.; Niyogi, P. Laplacian score for Feature Selection. *Adv. Neural Inf. Process. Syst.* **2005**, *18*, 507–514.
41. Zhao, Z.; Liu, H. Spectral feature selection for supervised and unsupervised learning. In Proceedings of the 24th International Conference on Machine Learning, New York, NY, USA, 20–24 June 2007; pp. 1151–1157.
42. Solorio-Fernández, S.; Martínez-Trinidad, J.F.; Carrasco-Ochoa, J.A. A new unsupervised spectral feature selection method for mixed data: A filter approach. *Pattern Recognit.* **2017**, *72*, 314–326. [CrossRef]
43. Paniri, M.; Dowlatshahi, M.B.; Nezamabadi-Pour, H. MLACO: A multi-label feature selection algorithm based on ant colony optimization. *Knowl.-Based Syst.* **2020**, *192*, 105285. [CrossRef]
44. Hashemi, A.; Dowlatshahi, M.B.; Nezamabadi-pour, H. Ensemble of feature selection algorithms: A multi-criteria decision-making approach. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 49–69. [CrossRef]
45. Hashemi, A.; Dowlatshahi, M.B.; Nezamabadi-pour, H. A pareto-based ensemble of feature selection algorithms. *Expert Syst. Appl.* **2021**, *180*, 115130. [CrossRef]
46. Raquel, C.R.; Naval Jr, P.C. An effective use of crowding distance in multiobjective particle swarm optimization. In Proceedings of the Annual Conference on Genetic and Evolutionary Computation, Washington, DC, USA, 26 June 2005; pp. 257–264.
47. Frank, A.; Asuncion, A. UCI Machine Learning Repository. Available online: https://archive.ics.uci.edu/ml/index.php (accessed on 15 January 2022).
48. Kaggle: Your Machine Learning and Data Science Community. Available online: https://www.kaggle.com/ (accessed on 15 January 2022).
49. Caravan Insurance Challenge-Coil Challenge 2000. Available online: https://www.kaggle.com/uciml/caravan-insurance-challenge (accessed on 15 January 2022).

50. Huang, Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Min. Knowl. Discov.* **1998**, *2*, 283–304. [CrossRef]
51. Lovász, L.; Plummer, M.D. *Matching Theory*; American Mathematical Soc.: Providence, RI, USA, 2009; Volume 367.
52. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley Series in Telecommunications and Signal Processing; Wiley: Hoboken, NJ, USA, 2006.