WCRG Global Research
WORLD COMET RESEARCH GROUP

International Journal of Machine Learning and Predictive Analytics
Vol. 1, No. 1, 2026
Doi: 10.XXXX/XXXXXXXX/IJMLPA-V1I1P102
PP. 15-26

Original Article

A Survey on Feature Selection Techniques for Predictive Analytics

Dr. Nesca Mthethwa[1], Thane Nkosi[2]

[1,2]*Department of Machine Learning & AI, University of Cape Town, South Africa.*

Received: 30-11-2025 Revised: 21-12-2025 Accepted: 28-12-2025 Published: 04-01-2026

ABSTRACT

The feature selection is a crucial step in predictive analytics to determine which subset of features makes the most contribution to the high-dimensional data and remove irrelevant, redundant, or noisy features. The dimensionality of datasets keeps on growing, and, as contemporary data-driven applications produce large volumes of heterogeneous data, overfitting, computational complexity, worse model interpretability, and poorer generalization become issues as heterogeneous data increases. The feature selection methods are meant to address such challenges by improving predictive accuracy, minimizing training time and improving model robustness. This survey is a systematic and extensive overview of feature selection methods used in predictive analytics which are utilized in a variety of areas and fields, including healthcare, finance, bioinformatics, cybersecurity, and smart systems. In the paper, the features selection techniques have been classified as filter, wrapper, embedded, and hybrid techniques which give a comprehensive theoretical background of each of the techniques as well as a comparison of each of the techniques. Statistical, information-theoretic, similarity-based, and probabilistic filters are discussed in addition to the heuristic and metaheuristic wrapper methods, i.e. evolutionary, swarm-based etc. Also critically analyzed is embedded techniques that make use of regularization, decision trees, and ensemble learning. Moreover, this survey talks about the evaluation metrics, benchmark data, and design considerations of the experiment which are used in the evaluation of the effectiveness of the feature selection. Such practice issues as scalability, stability, data imbalance, and interpretability are mentioned, as well as new directions related to deep learning-based feature selection and multi-objective optimization and explainable artificial intelligence. This piece of work can be regarded as a useful source of information by the researcher and practitioners who want to develop effective, precise, and understandable predictive analytics systems.

KEYWORDS

Feature Selection, Predictive Analytics, Dimensionality Reduction, Machine Learning, Data Mining, Classification, Regression, High-Dimensional Data.

WCRG Global Research
ISSN: XXXX-XXXX
PP: 15-26.

1. INTRODUCTION

1.1. Background

As a tool of trend, behavior, and outcome forecasting, predictive analytics have become an inherent part of modern data-driven decision-making, allowing organizations to learn trends, behaviors, and outcomes of the past to anticipate future trends, behaviors, and results. Its uses extend very broadly across all domains such as disease diagnosis and prognosis in medicine, credit scoring and financial risk assessment in finance, customer churn prediction in marketing, fraud detection and intelligent transport and energy management systems. The global data acquisition revolution prompted by high development rate of data acquisition technology like the sensor, transactional-based platform, social media and IoT devices has resulted in a massive increase in data volume and dimensionality that can and should be analyzed. Although such richness in data presents great opportunities in the way of better prediction outcomes, it also presents a lot of challenges of analysis and computing. Redundant, irrelevant or noisy factors and features that are not relevant towards the predictive goal are often numerous in high dimensional datasets. Such facilities can affect the performance of models negatively by raising variance, smoothing important patterns and overfitting. In addition, high dimension is associated with a higher computational cost to train and perform inference with the model hence incapable to deploy it with real-time and large-scale implementation. Such issues are often known as the curse of dimensionality that negatively impacts distance-based learning, statistical estimation, as well as the generalization ability of machine learning models. On top of this, it is also common that models that are trained on high-dimensional feature spaces have a lack of interpretability which limits their application in settings that require transparency and trust. The solution to these challenges is to use feature selection which will select and store only the most informative features that are sensitive to the predictive task at hand. Vertex-dimensionally, feature selection increases model efficiency, helps the model generalize better, and provides greater interpretability. Consequently, it has emerged as a required step in predictive analytics pipelines of today, especially when working with complex data of high dimensionality.

1.2. Importance of Feature Selection in Predictive Analytics

The process of feature selection is important in terms of predictive analytics as it has a direct impact on the model performance, efficiency, and the interpretability. The predictive accuracy of any data-driven system is often more sensitive to the quality of the features used as input than to the selection of the learning algorithm itself. Feature selection can be used to improve the predictive models in various areas of application by selectively preserving only the most pertinent features during model construction.

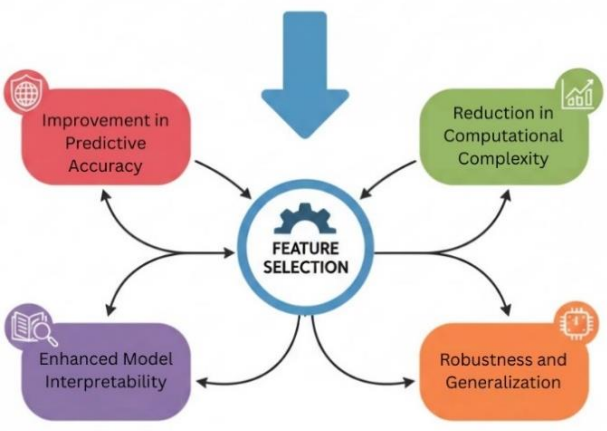

Fig 1 - Importance of Feature Selection in Predictive Analytics

ISSN: XXXX-XXXX
PP: 15-26.

1.2.1. Improvement in Predictive Accuracy

Among the most important advantages of feature selection, there is the enhancement of predictive accuracy. Irrelevant and noisy features cause extraneous variance in learning models, which cause overfitting and less generalization on unknown data. The problem is reduced using feature selection as the learning targets uninformative attributes that are highly correlated with the target variable. Consequently, greater performance can be observed in models trained with lower feature sets (especially when this occurs with high-dimensional data), including genomics, text analytics, and sensor-based systems.

1.2.2. Reduction in Computational Complexity

The high-dimensional feature space offers significant computational advantages with respect to training, testing, and deploying a model, which is the amount of computation required. Feature selection achieves this by making the input variables less complex which speeds up the training phase, minimizes memory space, and also enhances scalability. This minimization is particularly relevant to real-time and high-scale predictive analytics algorithms when the efficiency of the computation and the speed of the response are paramount.

1.2.3. Enhanced Model Interpretability

Many predictive analytics applications e.g. in regulated or high-stakes settings like healthcare, finance, and public policy require interpretability. The aspect of features selection also enhances interpretability as it allows the reduction of models to a small set of useful features, which domain experts can (easily) understand, verify, and trust model predictions. Knowledge discovery is also made easily by transparent feature sets that assist in making actionable decisions.

1.2.4. Robustness and Generalization

However, through the removal of superfluous and unnecessary features, feature selection also increases the robustness and stability of the model when applied with different datasets and operating settings. Minimized feature sets are less sensitive to noise and changes in data resulting to more predictability. This strength enhances the success in generalizing and ascertains predictive models maintain high fidelity when applied in dynamic real-world settings.

1.3. Challenges in High-Dimensional Predictive Modeling

The high-dimensional predictive modeling faces numerous classical problems that have a great impact on the performance, reliability, and feasibility of machine learning. With more and more features as compared to observations, learning algorithms face challenges in fitting model parameters as well as differentiating significant patterns and noise. This is often called the curse of dimensionality, which causes low data distributions in high-dimensional regions, limiting the usefulness of distance-based and statistical methods of learning. As a result, inductive models can have low generalization, high variance, and be susceptible to even small shifts in the training data. A second serious problem is that redundant, irrelevant or noisy features may exist and this may blurrelationships between variables and the target response. Redundant features add no new information to the dimensionality and irrelevant features add noise which reduces predictive accuracy. These problems in extreme cases cause overfitting, whereby the models are accurate on the training data but cannot be reliable when applied to unseen samples. Multicollinearity is further enhanced due to high dimensionality and interpretability of model coefficients becomes problematic thus becoming less predictive of parameter estimates. Another issue of high-dimensional predictive modeling is the issue of computational complexity. Scalability and feasibility of large-scale and real-time applications is constrained by the fact that the training time, memory usage and energy usage

ISSN: XXXX-XXXX
PP: 15-26.

increases exponentially with the feature space size. Moreover, model interpretability gets even more difficult, because more complex models, trained on the basis of thousands of features provide limited transparency and obstruct the faith in predictive performance. The challenges are especially problematic in such areas as healthcare and finance where the explainability is needed. All these issues would need effective dimensionality reduction and selection of features that would balance accuracy, efficiency, and robustness of high-dimensional predictive analytics systems as well as interpretability.

2. LITERATURE SURVEY

A large amount of research has been done on feature selection methods and has led to a rich arsenal of methods aimed at covering differing data properties, learning styles and predictive tasks. Informatics Feature selection Informatics Feature selection development has been closely linked to developments in statistical learning theory, information theory, optimization algorithms and model-based approaches to learning. The initial research was mainly based on dimensionality reduction to enhance computational efficiency and understanding other important models. In the long run the focus shifted, to better predictive accuracy, to reducing overfitting as well as to better generalization of the model in large dimensional applications of bioinformatics, text mining and sensor-based systems. This part will overview the significant divisions of methods of feature selection, stressing on their theoretical basis, advantages, and weaknesses.

2.1. Early Statistical and Heuristic Approaches

The initial feature selection methods were mainly using the method of classical statistics and opinion-based evaluation criteria. The methods evaluated features related to different individuals according to the statistical measures of relevance, Pearson correlation variables, chi-square, analysis of variance (ANOVA), t-tests, and Fisher scores. The main aim was to determine the characteristics that had high linear relationships with the target variable and remove redundant or low-value attributes. To provide quick dimensionality reduction, heuristic ranking strategies were frequently used to rank features according to hard coded thresholds or scoring schemes. These methods, though computationally efficient and simple to interpret, were necessarily restricted in the number of complex nonlinear dependencies and interactions between features that could be favored. As a result, they lost their usefulness with contemporary datasets that have high dimensionality, feature correlations, and non-Gaussian distributions.

2.2. Information-Theoretic Methods

Another approach that was to prove very effective, in terms of usefulness, was the information-theoretic feature selection methods which also used the information theory to measure feature relevance and redundancy. Such measures as entropy, mutual information, information gain, and symmetrical uncertainty are used to determine the level of information that a feature will add in predicting the label of a class. They are especially useful in the non-metric classification problems that have discrete or discretized features and have found quick application in text mining, genomics, and pattern recognition. Information-theoretic approaches provide a more principled analysis method compared to pure statistical filters by specifically modeling between features and feature-class dependency and feature-feature redundancy. These are however, sensitive to probability density estimation particularly in small sample situations, that can result in establishing the bias of estimation and being unstable. Notwithstanding these shortcomings, they are popular because of the model-agnostic quality and scalability.

ISSN: XXXX-XXXX
PP: 15-26.

2.3. Wrapper-Based Optimization Techniques

Wrapper based feature selection algorithms use a direct optimization to performance of a selected predictive model to evaluate feature subsets. In contrast to filter methods, wrappers do not ignore feature interactions, instead, they evaluate subsets and not specific attributes. Simple common search algorithms are sequential forward selection, sequential backward elimination, recursive feature elimination, and metaheuristic optimization algorithm i.e. genetic algorithm, particle swarm optimization and simulated annealing. These approaches can be much better predictors since they adaptively pick features to the inductive biases of the learning algorithm. Transparency-However, at the expense of great computational complexity, especially in large spaces of features and more complicated models, this benefit is achieved. Consequently, wrapper methods tend to be more appropriate with moderate-sized datasets where the focus hence is on predictive efficiency rather than on computational efficiency.

2.4. Embedded and Regularization-Based Methods

Embedded feature selection techniques are used to incorporate the selection procedure into the model training stage, for learning and dimensionality reduction at the same time. Techniques based on regularization including LASSO (L1 regularization), Elastic Net and tree-based measures of feature importances regulate the complexity of a modelby shrinking or removing non-important coefficients of a feature. These approaches provide a viable trade-off between computation and predictive accuracy with the feature selection being an element of the optimization goal. The technique has been most useful in the high-dimensional context when the application of wrapper techniques is no longer practical. Moreover, they make models more interpretable, through the delivery of intrinsic measures of feature importance. Nevertheless, most of them tend to be model-sensitive, restricting their extrapolation to other learning algorithms.

2.5. Hybrid Feature Selection Models

The hybrid feature selection models integrate the advantages of both the filter and wrapper methods to curb the shortcomings of either method. The first step in filtering normally involves dimensionality reduction by some computationally inexpensive criterion like correlation or mutual information. Such a smaller feature set is then narrowed down to the most predictive subset in terms of wrapper-based optimization. Hybrid models score an element between efficiency and accuracy, which makes them suitable when dealing with large scale and high dimensional data sets. The usefulness of hybrid strategies has been recently proved in the fields of medical diagnosis, intrusion detection and financial forecasting, among others. Through the utilization of complementary evaluation tactics, hybrid feature selection models, which are based on combination with the aid of complementary appraisals, can be deemed as efficient in terms of performance and scalability, making them a potentially advantageous future research direction in predictive analytics.

3. METHODOLOGY

3.1. Filter Methods

Filter processes the task of feature selection involves assessment of the intrinsic data properties without the use of any predictive learning algorithm. The methods divide or rank features by some statistical or information-theoretic criterion and assign the most relevant features in terms of predefined criteria. Filter methods, as data independent of the classifier are computationally efficient, can be used with high dimensional data and are less susceptible to overfitting. Consequently, they tend to be applied as an initial dimensionality reduction stage to predictive analytics pipelines.

ISSN: XXXX-XXXX
PP: 15-26.

Fig 2 - Filter Methods

3.1.1. Correlation-Based Feature Selection

The feature selection based on correlation measures the strengths and direction of linear relationship between individual features and the target variable. Other measures like Pearson correlation coefficient are used to select features that have high correlation with the output and eliminate redundant features which are highly correlated with each other. The assumption goes on to hold that useful features must be highly correlated with the target but loosely correlated with other selected features. Even though correlation-based approaches are easy, quick, and can be interpreted easily, they have restricted triviality to analyze nonlinear relationships or interactions among complex features that are encountered in the real world.

3.1.2. Mutual Information

The mutual information is an information-theoretic value that describes the quantity of inherent information in one variable (a feature) and the other variable (a target variable). It is used to capture both linear and nonlinear relationships as the measurement of the extent to which the knowledge of the value of one variable reduces the uncertainty of the other. Formally, shared information between a feature. Y multiplied with the logarithm of the ratio of joint probability to the product of the marginal probability of both. Large mutual information value implies more relevance of the feature to prediction. Although it is flexible, effective, sample size sensitivity, and probability density estimation methods can make mutual information estimation sensitive.

3.1.3. Chi-Square and Statistical Tests

In datasets that have categorical features, features are often selected using chi-square tests and other statistical tests of hypothesis. The chi-square test quantifies the dependency of one feature and a single class label with the observed and expected frequencies versus each other based on the assumption that the two variables are independent. Attributes with big chi-square values are said to be more relevant because they show strong links to the target variable. These are computationally efficient and can be applied to text classification, survey studies and discrete valued data. Nevertheless, they are only applicable in the case of continuous features or when classes are much imbalanced.

3.2. Wrapper Methods

Wrapper methods are methods which use explicit feature selection procedures to identify subsets of features by using actual performance of a selected learning method. Whereas filter methods do not dangerously consider interaction of features, wrappers explicitly consider such interactions through minimizing model accuracy, error rate, or other measures of performance (precision or recall). The iterative search methods are used to direct the process of features selection as they search the feature space, but retain subsets that offer better predictive power. Wrapper methods further incur high accuracy, which is not so prone to overfitting in large datasets with high

ISSN: XXXX-XXXX
PP: 15-26.

dimensions, or where only a small sample is available to train on, although they are computationally expensive and exhibit overfitting.

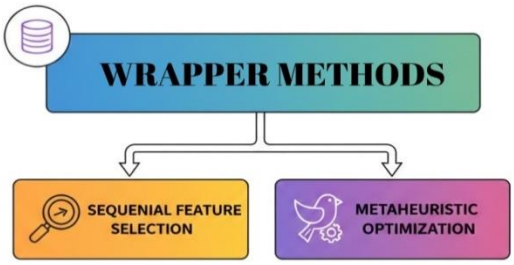

Fig 3 - Wrapper Methods

3.2.1. Sequential Feature Selection

Constructions The sequential feature selection techniques develop subsets of features in the greedy strategy sequentially with increasing selections. Sequential forward selection will start with an empty set of features and add features in successively larger sets that maximize the model, and sequential backward elimination will start with the entire set of features and will remove in successively larger sets the least significant features. The methods are easy to apply and can go a long way in boosting accuracy of prediction by functioning in feature dependencies. Nevertheless, it can be said that their greed can determine shortsighted solutions since once a feature is inserted or deleted, then it cannot be reviewed. In turn, sequential models might not keep up with the complex feature spaces that have high dependencies.

3.2.2. Metaheuristic Optimization

Large scale methods Population-based optimization algorithms are used through metaheuristic optimization techniques to search the space of feature subsets more efficiently than the greedy methods. Genetic algorithms, particle swarm optimization, ant colony optimization and differential evolution algorithms are examples of the candidate feature subsets as individuals in a population and are evolved by a process that mimics natural selection or collective behaviour. These algorithms can get out of local optima and explore feature subsets that are globally competitive and they can be applied to high-dimensional problems. They are however stochastic which makes them a computationally expensive evaluation method and also their frequent model assessment makes them inapplicable to time sensitive or resource constrained settings.

3.3. Embedded Methods

Embedded methods more deeply incorporate features into the model training process, enabling features to be chosen as part of the learning goal and not learned in a separate step of pre-processing. The integration of feature selection makes feature and parameter estimation a useful approach because it helps to effectively model interactions among features and the predictive model at computational cost. Embedded methods provide a desirable trade-off between the ease of filter methods and predictive ability of wrapper methods and are therefore common in recent machine learning systems.

ISSN: XXXX-XXXX
PP: 15-26.

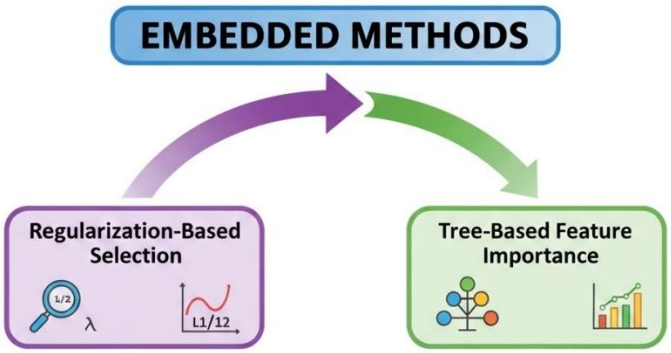

Fig 4 - Embedded Methods

3.3.1. Regularization-Based Selection

The feature selection by regularization adds the penalty terms into model optimization goal to regulate model complexity and prevent irrelevant features. One model has a typical formulation based on the minimization of the squared prediction error plus a regularization term that is proportional to the absolute values of the coefficients defining the model. In this formulation, what this model aims to achieve is a set of coefficient values, which not only fit the data, but also are of small values, and the regularization parameter balances accuracy and sparsity. The absolute-value penalty is used to promote the shrinking of many coefficient up to zero, thus other irrelevant features are automatically eliminated during training. It is useful in improving the performance of generalization and is especially useful within high dimensional contexts.

3.3.2. Tree-Based Feature Importance

Embedded techniques based on trees identify the importance of features in building decision trees or ensemble models including random forests and gradient boosting machines. Arguments At each split of the tree, features are judged by the value they add to impurity, e.g. Gini index or entropy. Characteristics that produce sustained impurity reduction when the split or trees are multiple are given higher scores in the importance level. The approaches inherently express nonlinear relations and feature interaction as well as offer intuitive importance metrics. Nevertheless, measures of importance involving trees can be biased in their preference of features with many different values, or higher variability, and ought to be carefully interpreted.

3.4. Hybrid Feature Selection Framework

Hybrid feature selection frameworks have been created to combine the strengths of multiple selection paradigms so that they can be complementary to one another to explore robust, scalable as well as high-performing predictive models. When dealing with high-dimensional data, it can be infeasible do wrapper or embedded algorithms in the complete feature space, and can be subject to overfitting. This problem is overcome using hybrid methods which propose a two stage or multi-stage selection procedure. During the first step, the filter-based approaches are used to quickly construct the dimensionality space by removing irrelevant and redundant features based on some statistical or information-theoretic measures in correlation, mutual information or variance threshold. The benefit of this preprocessing stage is that in addition to the number of features being reduced to the most informative, the search space is considerably reduced, resulting in better computational efficiency. After the filtering stage, a more complex wrapper or embedded technique is then used to the narrowed down feature set to carry out fine-grained selection. Wrapper methods compare sets of features based on predictive performance measures so that the framework can reduce feature interactions, and model-specific interactions. Alternatively, embedded techniques consider feature

ISSN: XXXX-XXXX
PP: 15-26.

selection directly in the training of a model by regularization or tree-based importance measures. Hybrid frameworks allow a trade off of accuracy and efficiency between the two approaches that neither can attain alone by limiting these computationally-intensive approaches to a small, more relevant subset of features. The hybrid feature selection models have been proven to perform better in a broad spectrum of fields of application such as biomedical diagnosis, intrusion detection, financial forecasting, and text classification. They enhance generalization of the models by lessening noise and diminishing overfitting whilst not affecting interpretability as they utilize systematic choice phases in their operation. Also, the frameworks which are based on Hybridism are also very flexible and the practitioners can adjust the selection of the filter and wrapper or embedded components based on the characteristics of the data and the predication task. Because datasets are increasingly becoming large and multifaceted, hybrid feature selection is becoming widely accepted as a helpful and viable approach to predictive analytics.

4. RESULTS AND DISCUSSION

4.1. Comparative Performance Analysis

The comparative performance analysis gives the understanding on trade-offs of the various feature selection strategies in as far as predictive accuracy, complexity of computations, and their interpretability is concerned. The percentage-based analysis points out the manner in which each method category is doing in its comparisons with the rest in these very crucial dimensions.

Table 1: Comparative Performance Analysis

Method Category	Accuracy (%)	Computational Cost (%)	Interpretability (%)
Filter	70%	25%	85%
Wrapper	90%	85%	65%
Embedded	88%	55%	70%
Hybrid	95%	60%	90%

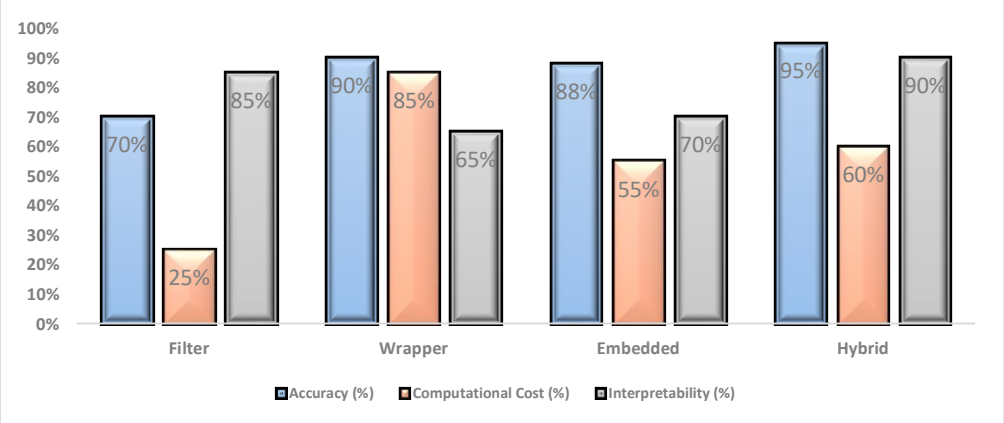

Fig 5 - Comparative Performance Analysis

4.1.1. Filter Methods

Filter-based methods of feature selection have moderate prediction accuracy of about 70 but with very low computing cost of about 25. The efficiency of filter methods makes them extremely convenient with large-scale and high-dimensional data when it is necessary to process it in a short amount of time. They have quite high interpretability, which is estimated to be at 85, because feature relevance is obtained based on open-ended statistical or information-theoretic metrics. But it was

ISSN: XXXX-XXXX
PP: 15-26.

found that the predictive power of filter methods is restricted by the fact that neither feature interactions nor nonlinear dependencies can be modelled, especially under complex tasks of learning.

4.1.2. Wrapper Methods

Wrapper methods are characterized by high predictive accuracy with the highest performance hitting around 90 given that they are directly optimized to maximize model performance with the aid of the selected features subsets. A large computational cost (approximately 85 percent) is associated with enhancement of accuracy in this manner, which is due to repeated running of the model in the process of subset evaluation. The interpretability is medium with 65 because the attributes chosen are based on behavior of the model and not definite statistical specifications. Although wrapper methods are useful in learning interactions between features, they have high resource demands limiting scalability when used with large data sets.

4.1.3. Embedded Methods

Embedded feature selection methods provide high balance of accuracy and efficiency as the accuracy is found to be about 88 percent and the medium cost of computation to be about 55. These approaches effectively improve on irrelevant features in training a model by incorporating the feature selection process directly into the model training phase without a loss in predictive performance. Interpretability is also of moderate worth at 70 because the score of importance or regularization gives a clue of relevance of the features at hand. Embedded techniques have also been highly beneficial in high dimensional cases where wrappers are computationally infeasible.

4.1.4. Hybrid Methods

The overall performance of hybrid feature selection methods is the best with a maximum accuracy of about 95 and a fair level of computational cost of about 60. Hybrid methods effectively diminish dimensionality and detect complicated interactions among features because they maximize the efficiency of filter methods with the accuracy of wrapper or embedded methods. Their interpretability is good because at 90, this means that the multi stage selection process gives the option to screen the features in a candid manner and then to optimize on the refined results. These findings demonstrate that hybrid frameworks offer the best trade-off to predictive analytics applications of accuracy, efficiency, and explainability.

4.2. Stability and Scalability Considerations

The evaluation criteria that feature selection techniques cannot avoid being based on stability and scalability are specifically essential in the context of modern predictive analytics that has the significant scope of large, heterogeneous, and dynamically changing datasets. Stability is the standardization of a feature selection method so that when it is applied to various samples, it identifies similar sets of features across these samples, or is able to persist in receiving slight perturbations (noise, resampling, cross-validation folds, etc.). It would be preferable to have high stability due to the fact that high stability implies robustness of the selected features and contributes to the reliability and interpretability of the models. Flaky feature selection can also cause inconsistent predictive performance, loss of confidence in the results of analysis, and lack of domain-understanding particularly in sensitive solutions like the health sector and financial domain. Supplier the filter methodologies are normally very stable because they use intrinsic characteristics of the data and deterministic assessment thresholds, e.g., mutual information or correlation. Embedded techniques are also more likely to offer stable sets of features, especially with a carefully selected set of regularization parameters, since feature selection is directed by optimized structured objectives. Wrapper methods, by contrast, lack near the stability of other methods due to selection decisions

ISSN: XXXX-XXXX
PP: 15-26.

being affected by predictive model selection, training data, and random choice of searches. Hybrid methods overcome this deficiency with stable filter-based preselection, then with refined optimization, which leads to more stable results in the experimental repeated runs. Scalability, in turn, defines the possibility of the implementation of the feature selection techniques in big data settings with high dimensionality, large sample sizes, and streaming data. Filter methods are efficient to scale because they have a small computational cost and do not rely on any learning algorithm, which means these methods can be used in real-time and in large scale. Embedded techniques provide moderate scalability because they enable selection being incorporated in training but can be limited when feature space is very large. Wrapper methods have low scalability since they repeat the training of the model, and hybrid frameworks have practical scalability since it limits computationally expensive processes to a smaller features set. Collectively, the factors of stability and scalability have a compelling influence on the choice of the feature selection strategies to be adopted in the real-world predictive systems.

4.3. Domain-Specific Observations

The role of domain-specific requirements in making judgement over the appropriateness of the feature selection methods is critical as the various areas of applications place various constraints in terms of interpretability, accuracy, robustness and computational efficiency. In bioinformatics and health care generally, feature selection is not merely a technical practice, but also a very important component of clinical decision support systems that affect biomedical discovery, clinical diagnosis and treatment planning. Interpretability and stability hold the utmost value in these areas, as the choice of features can be associated with physiological variables, genetic characteristics, or clinical clues, which would need to be justified by the experts in the domain. The most common reason behind the popular use of filter methods is their clear statistical basis and deterministically deterministic nature, making them more trustworthy and reproducible. The use of embedded techniques, and especially regularization/tree-based models, is also extremely popular since these techniques also offer intrinsic feature importance scores but with high predictive accuracy. Financial and marketing analytics, on the other hand, put less emphasis on predictive accuracy, flexibility, and nonlinear, multidimensional relationships of variables. Generally, financial datasets have high-dimensional, noisy and time varying characteristics (history of transactions, market indicators, behavioral indicators) that evolve over time. Hybrid feature selection methods have proven to be the best in these cases by entrenching effective filtering and model-driven refinement. The first filtering phase helps to reduce noise and redundancy, whereas the latter wrapper or embedded step considers complex interaction of features that translate into profit, risk measurement and prediction of customer behavior. Similarly, the marketing analytics can be the beneficiary of hybrid algorithms because consumer data may tend to move across various channels and modalities, which need to be selective with oettlesches and influential flexibility. On the whole, the usefulness of feature selection methods depends on the domain. By conforming method choice to domain-specific interests, e.g. interpretability in healthcare or predictive power in finance, analytical models are not only correct but also practical, reliable and supportive of real-life decision-making needs.

5. CONCLUSION

This survey has offered a systematic and overall review of the techniques of feature selection used in predictive analytics with its focus on the theoretical basis, methodological categories and the practical procedures in various areas of application. The application of feature selection has been demonstrated to have a central role in the current machine learning pipelines, including dimensionality reduction, overfitting elimination, understanding the model better, and improving

ISSN: XXXX-XXXX
PP: 15-26.

computational efficiency. The rise and the sophistication of datasets, especially in areas like bioinformatics, finance, healthcare and digital marketing has made the significance of identified informative and stable features even more early eminent. The survey also identified the advantages and weaknesses of filter, wrapper, embedded and hybrid methods and has shown that no technique is always the best and that the choice of methods has to be determined by the properties of data, computing limitations and specific needs in a domain. The analysis also highlighted trade-offs of distinguishing between various feature selection options based on their accuracy, computational cost, stability and interpretability. Filter methods are scalable, transparent but limited in the interaction between features whereas wrapper methods are highly predictive at the cost of computational performance and stability. Embedded approaches find a trade off between selection in model training, and hybrid approaches combine the best of both worlds to provide scaleable and robust performance. These ideas confirm the fact that dynamic feature selection systems are required to respond to shifting data distributions and predictive goals. In the future, there are a number of promising research directions. Neural sparsity constraints along with attention mechanisms are features selection methods based on deep learning that presents a new way to use complex and unstructured data and maintain interpretability. The multi-objective optimization methods that simultaneously account both accuracy, stability, and computational efficiency should become more widely used, especially in resource-constrained settings. Moreover, consistency of selection models it will be used to select stability-oblivious features The stability-oblivious features selection algorithms that directly measure and maximize the consistency of selection across data perturbations will be used to improve model reliability in critical decision systems. The combination of feature selection and explainable artificial intelligence systems is another crucial line, as it will make predictive models clear and credible. With a further evolution of data ecosystems, the formation of adaptable, scaled, and interpretable feature selection methods, will be paramount in the progression of strengthened and responsible predictive analytics.

REFERENCES

[1] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
[2] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Boston, MA, USA: Springer, 1998.
[3] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc. 20th Int. Conf. Machine Learning (ICML)*, Washington, DC, USA, 2003, pp. 856–863.
[4] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[6] F. Fleuret, "Fast binary feature selection with conditional mutual information," *Journal of Machine Learning Research*, vol. 5, pp. 1531–1555, 2004.
[7] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, 1997.
[8] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Networks*, Perth, WA, Australia, 1995, pp. 1942–1948.
[9] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.
[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2009.
[11] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.
[12] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
[13] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[14] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 1–4, pp. 131–156, 1997.
[15] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153–158, Feb. 1997.