WCRG Global Research
WORLD COMET RESEARCH GROUP

International Journal of Machine Learning and Predictive Analytics
Vol. 6, No. 1, 2023
Doi: 10.67228/3142788X/IJMLPA-2023PI3R7H
PP. 01-14

Original Article

A Comparative Study of Supervised Learning Algorithms for High-Dimensional Data

Sanjay Verma[1], Naveen Kumar[2]

[1]*Operations Manager, HCL Technologies, India.*

[2]*Product Manager, Zoho Corporation, India.*

Received: 12-01-2022 Revised: 12-18-2022 Accepted: 12-28-2022 Published: 01-02-2023

ABSTRACT

High-dimensional data are now ubiquitous in the modern science and industry, such as bioinformatics, text mining, computer vision, finance, and cybersecurity. A prominent feature of such data is having many features in comparison with the number of observations, which may cause the judgement problem of the curse of dimensionality, greater computational cost, feature overlap, and overfitting. Though supervised learning algorithms are extensively used to do predictive modeling, they have very different performance properties in high dimensional feature space. The paper contains a thorough comparison of some of the most popular supervised learning algorithms in the high-dimensional data analysis scenario. The paper provides a systematic comparison between the linear, non-linear, probabilistic, and ensemble-based classifiers, which are: Logistic Regression, Support Vector Machine, k -Nearest Neighbor, Decision Tree, Random Forest, Naive Bayes, and Artificial Neural Network. Special attention is given to the study of the behaviour of an algorithm based on scalability, ability to generalize, resistance to noise, feature sparsity, and interpretability. Besides, the paper explores how dimensionality reduction and feature selection methods impact on the performance of classification. It suggests a single experimental procedure with standardized preprocessing pipelines, cross-validation schemes and performance metrics accuracy, precision, recall, F1-score and cost of the computation. To give the concept theoretical background, mathematical formulations of learning objectives and decision functions are given. The comparative analysis indicates that there is no universal algorithm that has the best performance in all high-dimensional conditions; the performance highly depends on the sample size, the features correlation, the level of data distribution as well as noise. This study has practical implications on researchers and practitioners to consider the proper supervised learning model to use the high-dimensional datasets and identifies future research opportunities in scalable and interpretable learning.

KEYWORDS

High-Dimensional Data, Supervised Learning, Classification Algorithms, Feature Selection, Curse Of Dimensionality, Machine Learning Evaluation.

ISSN: 3142-788X
PP: 01-14.

1. INTRODUCTION

1.1. Background

The recent development of data acquisition, data storage and sensing technologies has instigated an exponential augmentation in the dimensionality of contemporary datasets, and numerous genuine world-appropriate activities now entail thousands or even millions of qualities. Examples are notably gene expression analysis where biological samples are described in terms of the activity levels of tens of thousands of genes, text mining systems actually represented in terms of high-dimensional and sparse term-frequency vectors, and image analysis systems actually represented in terms of high-resolution pixel values or deep feature embeddings learned by convolutional neural networks. It, though, provides these high-dimensional representations with providing better descriptive richness and might have an ability to describe complex underlying patterns, they also pose a high-computational and statistical burden to existing machine learning algorithms. Guided learning methods which rely on labeled training data to build predictive models have met with significant success in a wide variety of fields including bioinformatics, natural language processing and computer vision. Nevertheless, their accuracy tends to decline due to the fact that the number of dimensions grows exponentially and that the significant statistical structure in the data is diluted. In high-dimensional spaces, similarity metrics that are based on distance are less discriminative, there is redundancy among features and overfitting tends to occur significantly and especially when there are a small number of provided labelled samples. The problems complicate the training of models, diminish the ability to generalize, as well as interpretability. As a result, systematic knowledge of how various supervised learning algorithms behave with high dimensional data has a pivotal role in efficient model selection, performance control and creation of scalescaleable learning methods that can meet the challenges of the application of high dimensional data.

1.2. Comparative Study of Supervised Learning Algorithms

Comparative learning algorithm is an important task that can be used to learn about the weaknesses, strengths, and the applicability of supervised algorithms in analyzing high-dimensional data. The concept of supervised learning covers an extremely broad scope of models such as linear classifiers, margin-based techniques, probabilistic, tree-based ensemble, and neural networks, with each using a different theoretical basis. When these algorithms are utilized with high-dimensional data sets they have different characteristics with regard to predictive accuracy, generalization capabilities, computational efficiency and interpretability. A systematic comparison enables the researchers and practitioners to determine which of the models work best with different data properties, including sparsity of features, sample size, and noisiness. Linear models, especially Logistic Regression, are commonly preferred in high-dimensional estimations because they are easy to use, interpretable, and effective at using regularization to suppress overfitting. The margin-based classifiers such as the Support Vector Machines improve generalization by maximizing the separation of the classes thus proving especially efficient in the sparse and high dimensional feature space. In contrast, distance-based methods, including the k-Nearest Neighbors, are more prone to deteriorated performance with the dimensionality, as the distance measures lose their discriminative ability. Probabilistic models, such as Naive Bayes, have a low cost in terms of computational resources and

ISSN: 3142-788X
PP: 01-14.

scalability, particularly where feature independence results are roughly met. Decision Trees and Random Forests are tree-based and ensemble methods which can easily capture the complex non-linear relationships and feature interactions. Nevertheless, their complexity in models can be associated with their higher costs in terms of computation and increased risk of overfitting in low-sample and high-dimensional conditions. Neural network-based methods are effective representation learning methods, with high representation learning properties, but they need great quantities of labeled data and hyperparameter balancing to stabilize performance. Hence, comparative evaluation of the algorithm of supervised learning is necessary to choose suitable models, trade accuracy and efficiency, and inform further research in the high-dimensional machine learning.

1.3. Characteristics of High-Dimensional Data

High-dimensional data sets have a few unique properties, which have a profound sensitivity on the operation and the efficiency of supervised learning algorithms. The knowledge of these properties is crucial to designing effective models, feature engineering, and picking algorithm.

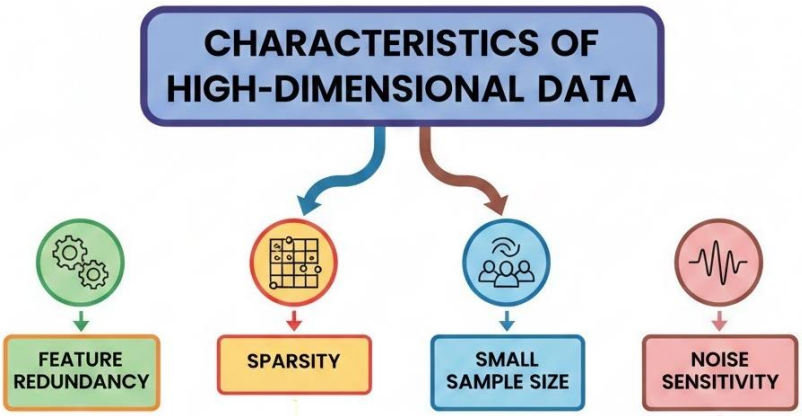

Fig 1- Characteristics of High-Dimensional Data

1.3.1. Feature Redundancy

A typical property of high-dimensional data is feature redundancy, in which there are several features that provide redundant or insignificant information. Such redundancy can theoretically be due to correlated variables, repeated measurements or generated features.

1.3.2. Sparsity

The case of having many feature values that have a value of zero or close to zero in the data set is referred to as sparsity. This is a well-known property in areas like text mining, recommender systems and bioinformatics. Although sparsity may decrease the storage requirements, it also creates problems with distance-based algorithms and similarity-based algorithms, as meaningful patterns are more difficult to identify. Sparsity exploitation models, including linear classifier and probabilistic models tend to be more efficient within these environments.

ISSN: 3142-788X
PP: 01-14.

1.3.3. Small Sample Size

In most cases of high-dimensional applications, the quantity of available training samples is considerably less than the quantity of features. Such an imbalance causes ill-posed learning problems, that is, several models can predict training data equally well, but not doing so on unseen samples. Small samples contribute to overfitting and decrease the statistical stability of acquired patterns and thus regularization and cross-validation are critical parts of the learning process.

1.3.4. Noise Sensitivity

High dimensions intensify the effect of depthfully noisy and irrelevant features that may distort the signal patterns and worsens a model. Noi sensitivity also amplifies variations in model predictions and makes the model less robust especially in the case of a complex model. Consequently, the challenges in noise processing by filtering features, regularizing data, and Hollywood active learning methods are essential to the accomplishment of trustworthy results in the analysis of high-dimensional data.

2. LITERATURE SURVEY

2.1. High-Dimensional Learning Challenges

The high dimensionality was found to be a problem in early investigations of statistical learning theory because of what is usually known as the curse of dimensionality. The dimensionality of the feature space grows exponentially with the number of features making available data samples sparse. This scarcity highly impairs the accuracy and consistency of a statistical estimation and distance-related measures of similarity. It was shown that, near high-dimensional spaces, the cyclone between similar and distant examples simplifies, and the conventional distance metrics, including the Euclidean distance, offer diminished information. Subsequently, proximity-based algorithms, such as k-nearest neighbors and kernel density estimation algorithms, have a significant drop in performance. Moreover, large dimensionality aggravates the computational complexity and overfitting, which have led to the creation of dimensionality reduction methods, feature selection methods, and regularization methods to enhance learning in case of high dimensionality.

2.2. Linear Models in High Dimensions

The linear models are extensively discussed in high-dimensional data with their mathematical simplicity and interpretability as well as efficient computation. Classical linear classifiers like Logistic Regression and Linear Discriminant analysis are problematic in cases where the number of features is greater than the number of samples thus resulting in ill-posed optimization problems. To resolve this concern, regularization methods of L1 (Lasso) and L2(Ridge) penalty were proposed to regulate model complexity and stabilize learning of parameters. L1 regularization encourages feature coefficients to become sparse and thus performs feature selection whereas L2 regularization encourages large weights and has enhanced numerical stability. Empirical experiments have repeatedly shown that regularized linear models are a competitively good model in high-dimensional environments, especially with the underlying data-generating mechanism being more or

ISSN: 3142-788X
PP: 01-14.

less linear or sparse, such that they become a formidable baseline model in a number of real-world problems.

2.3. Kernel-Based Methods

Support Vector Machines defined as kernel-based learners, especially SVMs, became popular to work with high-dimensional data based on their maximization of margins and because such learners do not take explicit representations of features. It has been discovered that in very high-dimensional settings like text classification and bioinformatics where feature space can have tens of thousands of dimensions, linear SVMs can be remarkably effective. Linear kernels are useful in such situations that can take advantage of the naturalities of sparse data and scale. On the other hand, non-linear kernelized SVMs can be used to detect complex decision boundaries but at high computational and memory costs with issue of increasing costs as the size of the dataset scale up. Literature points out that although there are strong theoretical guarantees of kernel methods, they are limited by scalability limits and ended up being limited as well by kernel selection issues in high-dimensional large scale applications.

2.4. Probabilistic and Bayesian Approaches

Naive Bayes, especially probabilistic models have gained a lot of use in high dimensional learning activities since they are simple and have sound theoretical basis. The naive Bayes is highly appropriate in areas that have many features and few training data due to the conditional independence assumption that removes much complexity in parameter estimation. Regardless of this robust assumption, empirical research shows that Naive Bayes can sometimes perform surprisingly quite well, in particular in text classification and document filtering problems that have sparse features and approximate independence. Bayesian methods also give some probabilistic explanations and uncertainty measure, which is useful in decision-related purposes. Nonetheless, they might become poor performers in the presence of the high dependency of features where hybrid models that lessen the independence assumptions hybrid models are encouraged.

2.5. Ensemble and Tree-Based Methods

Ensemble models like the Random Forests as well as Decision Trees have proven to be extensively used in case of high-dimensional data sets since they are able to capture both non-linear relationships and the interactions of features. Ensuring that there is generalization: the ensemble methods enhance this condition of generalization by combining many small learners and hence reducing variance, which advances strength. Nevertheless, the literature also states that tree-based models can be poorly modeled in high-dimensional (and low-sample-size) data, as they can easily overfit on features that are sporadic and not informative. An array of techniques such as random feature selection and pruning have been suggested to address this problem and their performance is sensitive to parameter tuning. Although ensemble techniques can be quite accurate in making predictions, predictive performance can be constrained by their interpretability and computational cost in high-dimensional large-scale applications.

ISSN: 3142-788X
PP: 01-14.

2.6. Neural Networks and Deep Learning

Neural networks and deep learning models are high representational capacity models, which asymmetrically represent non-linear trends in data. Nonetheless, they have a number of limitations when applied to other high-dimensional problems of learning, such as one having to use large labeled datasets to prevent overfitting. Studies show that with an increase in dimensionality, neural networks get susceptible to noise and irrelevant features, causing the neural network training to be unstable and the generalization to perform poorly. Successful deployment therefore requires good feature selection, dimensionality reduction and regularization methods like dropout and weight decay. Although deep learning has been shown to achieve a state of the art performance in areas with large amounts of data, it has been noted in literature that when dealing with small sample sizes and high-dimensionalities, simpler models are more likely to perform better than deep architectures because of superior biasvariance trade-offs.

3. METHODOLOGY

3.1. Problem Formulation

Assume a monitored learning dilemma founded on a cataloged sample size of N samples represented as:

$$\mathbf{D = \{ (x_i, y_i) \} \text{ for } i = 1 \text{ to } N,}$$

With all input vectors x i drawn to a d -dimensional real-valued feature space (Rd) and each label y i drawn to a finite set of C class labels {1, 2, …, C). The dimensionality d of feature space (d far much greater than N) is much larger than the available training samples N in high-dimensional learning situations. Such a disproportion of dimension in feature to sample presents considerable problems in feature-based model learning that affect overfitting risks, uncertainty in the estimation of model parameters, and low generalization. The main goal of the supervised learning in this scenario is to train a predictive model f, which predicts the functions of criteria of the input vectors in the high-dimensional feature space $\mathbb{R}$ present in the high-dimensional feature space $\mathbb{R}$ to the output space $\mathbb{R}$, in such a manner that the predicted outputs are close to the actual class labels. This learning process is usually modeled as an empirical risk minimization problem, the objective in which is to find a function f that minimizes the mean loss on the training set. This loss functional measures the difference between the estimated output f (x i) and the actual label y i of an example in the training dataset. Useful loss metrics are cross-entropy loss, used in classification problems, and hinge loss, used by margin-based classifiers. With the dimensionality of the environment, empirical risk may be minimized directly which tends to result in poor performance during generalization owing to the existence of multiple redundant or irrelevant features or noisy features. Therefore, further restrictions or regularization conditions are usually added to the optimization goal to restrain the complexity of the model and enhance robustness. These constraints permit models of less complexity, encourage feature selection sparsity, and alleviate the negative impact of dimensionality. Thus, the issue formulation is not merely concerned with the reduction of empirical risk but it also lays stress on balancing the model expressiveness and a generalization performance in high-dimensional feature spaces.

WCRG
Global Research
ISSN: 3142-788X
PP: 01-14.

3.2. Algorithms Considered

A combined variety of classical and modern classification methods is examined in order to assess the performance of supervised learning methods in high-dimensional data context. All algorithms are a variant of learning paradigms and have different advantages and disadvantages in terms of high feature dimensionality.

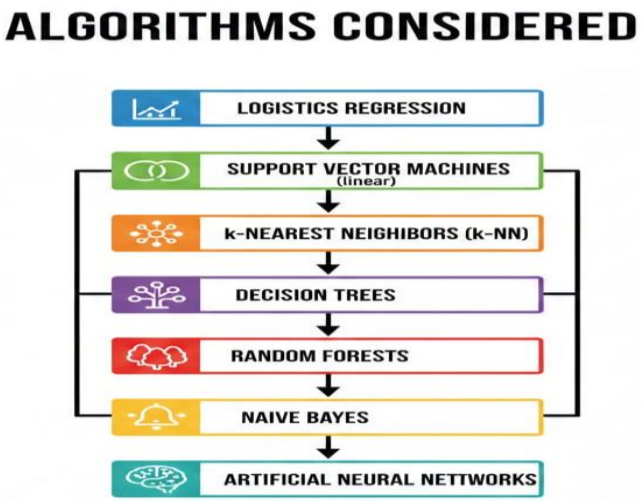

Fig 2 - Algorithms Considered

3.2.1. Logistic Regression

Logistic Regression is a popular linear classification algorithm that will compute the likelihood of the membership of a class based on a logistic function. High dimensional situations It is commonly used together with regularization methods to avoid overfitting as well as to enhance generalization. Since Logistic Regression is formulated as a convex optimization, it is computationally efficient, interpretable and can be used as a potent baseline to test more complicated models, especially when the relationship between features and class labels is in a more or less linear form.

3.2.2. Support Vector Machines (Linear)

The objective of the Linear Support Vector Machines is to find a hyperplane that maximizes the separation between data in multiple classes in the feature space. Their success in high-dimensional spaces is due to the fact that the optimization is based on a few points of the training referred to as the support vectors. Linear SVMs are especially applicable to sparse data and high dimensional data like text and genomic data and the linear SVMs have good generalisation as well as with comparatively low computational complexity than the non-linear kernel versions.

3.2.3. k-Nearest Neighbors (k-NN)

An example based learning algorithm is the k- Nearest Neighbors algorithm that classifies samples using the predominant sample among their nearest neighbors in the feature space. Simple and quite intuitive, k-NN is highly problematic in high-dimensional spaces where the meaning of distance metrics becomes low because of features sparsity. With a higher dimensionality, the

ISSN: 3142-788X
PP: 01-14.

discriminatory ability of the distance-based similarity measures diminishes in most cases, therefore, leading to poor classification and higher cost of computation.

3.2.4. Decision Trees

Decision Trees: Non-parametric models: Decision Trees are used to solve a classification problem by recursively subdividing the space of features in response to a feature threshold. They can capture non-linear interactions and they contain complexities. In large-dimensional data set however, overfitting is common with the use of Decision Trees particularly when the training samples are few. They need to be pruned properly and the depth of their pruning should be controlled to enhance their robustness and their generalization ability.

3.2.5. Random Forests

Random Forests are techniques of ensemble learning which build numerous decision trees based on bootstrap sampling and random selection of features. Random Forests decrease variance and enhance generalization by combining predictions of many trees as opposed to that of single Decision Trees. In spite of being more effective in high-dimensional data, their effectiveness is conditional on the proper parameter tuning and their computational complexity grows as the dimensionality of feature and the number of trees.

3.2.6. Naïve Bayes

Naive Bayes classifiers are computational probabilistic classifiers, which deal with the Bayesian theorem and assume that features are conditionally independent. This powerful assumption greatly eases the estimation of the parameters, rendering Naïve Bayes especially adequate in a high-dimensional and sparse data set. The algorithm is comparatively complex and may provide competitive performance on a bare-bones approach to classification, particularly of text and hinges more on features dependence in that context, though its relevance declines as the dependency increases.

3.2.7. Artificial Neural Networks

Artificial Neural Networks are neuron networks that have the ability to learn complicated non-linear mappings between input and output. They are highly-representative and need to be carefully-architectured, well-trained, and regularly-regularized to be useful in high-dimensional scenarios. Unless proper data or dimensionality reduction is performed, neural networks can be prone to overfitting and a feature selection and normalization must be considered an essential preprocessing task if a neural network is to deliver reliable results.

3.3. Mathematical Models

Formal analyses of the mathematical modeling of supervised learning algorithms can give insight into the nature of their optimization behavior and generalization features in high-dimensional spaces. In case of Logistic Regression, the learning problem is developed into a regularized empirical risk minimisation problem. The model aims at learning a weight vector w that reduces the total of

ISSN: 3142-788X
PP: 01-14.

logistic loss undoubtedly of every education sample. The logistic loss is used to compare the predicted archetype with the actual archetype by costing an error in classifying exponentially. High-dimensional averting overfitting: In high-dimensional cases, an L2 regularization is based on the claim that large weight components are penalized by including the squared Euclidean norm Route of the weight vector. The regularization parameter L regulates the trade-off between a minimized classification error and a simple model. This formulation leads to a convex optimization problem, which has a global minimum and this ensures that Logistic Regression is stable and efficient when the number of features is large. In Support Vector Machines (SVMs) one would aim to increase the size of the class spacing and give a tolerable number of errors in the decision making. SVM formulation is an expression of a minimum of two terms, the squared norm of the weight vector that imposes a maximum on the margin and the penalty term that is proportional to the sum of slack variables. These slack variables are used to measure the extent to which individual samples are not using the margin constraints. C is the regularization constant that finds the compromise between three maximization of the margin and the minimization of error of the training. The higher the value C is, the more the training samples are correctly classified, and the lower the value is, the greater is the margin and better the generalization. The accuracy of the training samples on the correct classification with a margin of at least one is subject to constraints, except that in the case of violations detected by slack variables. This formulation of constrained optimization allows SVMs to be effective in high dimensional feature spaces because the solution can rely on only a few critical samples, called support vectors. Thus, Logistic Regression and linear SVMs can be effectively used in learning problems that are of high-dimensionality because of their solid theoretical basis and robustness via regularization.

3.4. Evaluation Metrics

In order to fully evaluate the performance of the supervised learning algorithms in high dimensional data settings, various evaluation metrics are used. The metrics all represent a predictive accuracy, effectiveness by class, computational efficiency and model scalability.

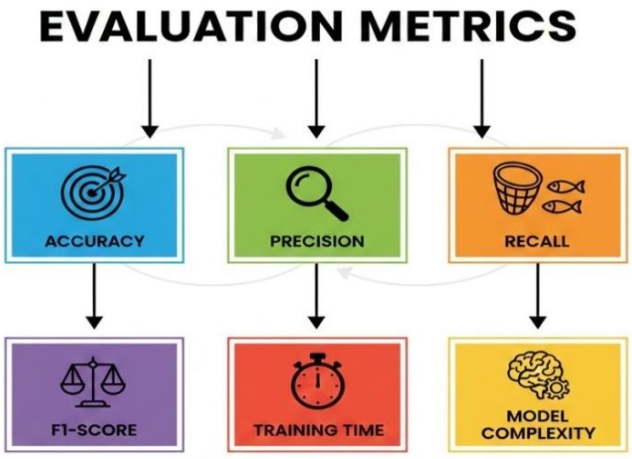

Fig 3 - Evaluation Metrics

ISSN: 3142-788X
PP: 01-14.

3.4.1. Accuracy

Accuracy is a measure of the number of accurately classified instances, as compared to the total number of samples. It gives a simple and easy to understand evaluation of the general model performance. Although the accuracy can prove practical to balanced data, in high dimensional classification programs with class imbalance, it can be a misleading factor since a high accuracy can be gained by over-training the majority type of class instead of gaining the influence of minor types of class patterns.

3.4.2. Precision

Precision is a measurement of the percentage of correct positive prediction compared to all predictions that are positively predicted. It indicates the consistency of favorable forecasting of the model. Where there is redundancy of features and noise which obscure features (high-dimensional context, machine learning) precision is especially valued, since applications that need a high score on predicted positive results (e.g. medical diagnosis or fraud detection) may demand great confidence in the prediction.

3.4.3. Recall

As we have noted previously, recall or sensitivity or true positive rate, is a metric, which measures how many real positive cases are made correctly by the model. The metric highlights the capability of a model to cover all the positives. Recall is important in high dimensional problems where it is important to lose positive examples at a serious cost since it assesses completeness of model prediction.

3.4.4. F1-Score

Harmonic mean of the precision and recall is known as the F1-score because it gives a balanced quality of a model. It comes in handy particularly when a disproportionate dataset is involved since it considers the false positive as well as the false negative. When the learning tasks are high-dimensional, the F1-score is more informative than the class-independent measure of accuracy because it uses a quality of the prediction of a specific class.

3.4.5. Training Time

Training time must be a computational cost indicator of the effort involved to learn the model parameters using the training data. The increase in the computational sphere of high-dimensional data can often make training time in such datasets a very important criteria measuring scalability and practicality. A smaller training time algorithm is better when a large number of samples need to be trained or with real-time requirements.

3.4.6. Model Complexity

The complexity of a given learning model can be described as model complexity, which is defined as the structural or computational complexity of a model, which can also depend on the number of parameters, depth, or ensemble size. Overfit and resource-heavy models are likely to

ISSN: 3142-788X
PP: 01-14.

occur in high-dimensional learning. Complexity assessment of models to achieve predictive performance/interpretability balancing, memory cost, and deployability is determined.

4. RESULTS AND DISCUSSION

4.1. Comparative Performance Analysis

The analysis of performance comparison measures the efficiency of supervised learning algorithms in terms of accuracy, F1-score, training time, and scalability in high-dimensional data home. The findings point to the trade off between predictive performance and computational efficiency.

Table 1: Comparative Performance Analysis

Algorithm	Accuracy (%)	F1-Score (%)	Training Time (%)	Scalability (%)
Logistic Regression	88	86	20	90
Linear SVM	92	88	45	80
k-Nearest Neighbors	72	70	85	40
Decision Tree	75	73	30	65
Random Forest	89	87	80	68
Naïve Bayes	70	68	10	92
Neural Network	90	88	95	70

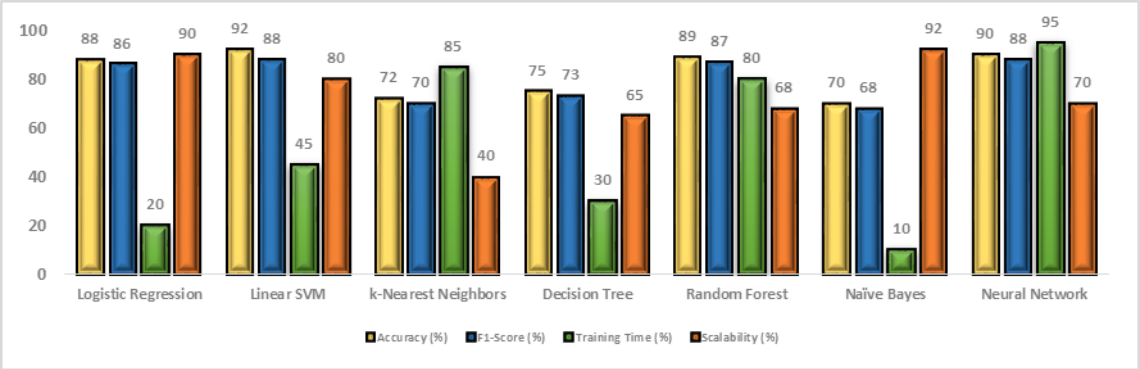

Fig 4 - Comparative Performance Analysis

4.1.1. Logistic Regression

Logistic Regression is also shown to show good overall performance of high accuracy and F1-score meaning that the high-dimensional settings can reliably classify. Its short training duration indicates high optimization and a low level of computation and high scalability ensures the applicability of the algorithm to large datasets. These properties make Logistic Regression a powerful baseline model in high dimensional learning.

4.1.2. Linear Support Vector Machine (SVM)

The highest accuracy of the models that have been compared is that of linear SVM which shows the effectiveness of this model in the separation of classes with the maximum margin in high-dimensional spaces. The medium increased training cost over the Logistic Regression is attributed to

ISSN: 3142-788X
PP: 01-14.

margin maximization and constraint. However, Linear SVM is an excellent option because of its good scalability and high F1-score in case it is used in the application with a high level of predictive accuracy.

4.1.3. k-Nearest Neighbors (k-NN)

k-NN shows medium accuracy and F1-score, which shows the failure of distance-based approaches in high-dimensional feature space. The expensive distance computations when inference is performed make its high training time a huge influence on scaling. When the number of dimensions grows, the generalization abilities of distance measures start to get worse, which restricts the use of k-NN in this case.

4.1.4. Decision Tree

Decision Trees have average performance regarding accuracy and F1-score with an advantage of non-linear interaction modeling of the features. They have relatively low training time, which implies that they are computationally efficient when constructing the model. Scalability is however average, where the high dimensional data raises the chances of overfitting and depths and pruning need to be controlled.

4.1.5. Random Forest

The high accuracy and F1-score are achieved by summing up predictions of several decision trees in the Random Forests, which effectively minimizes variance. This comes at an expense of takeover time as a result of building ensembles. Random Forests can compute high-dimensional data more effectively than single trees whereas their scalability is limited to the requirements of computational and memory.

4.1.6. Naïve Bayes

Naive bayes is easier and much more scalable to train because the assumptions of probabilistic formulation and independence are easy to set. It offers good baseline performance, especially in the case of sparse and high-dimensional datasets despite moderate accuracy and F1-score. Its efficiency allows its application in real-time and in large scale applications where computational resource is scarce.

4.1.7. Neural Networks

Neural Networks give high accuracy and F1-score, which indicates their capability in non-linear pattern of capture. Nonetheless, they require a very high training time, which points to high computational requirements, particularly when using high-dimensional feature spaces. Scalability is medium, but the quality of performance is greatly reliant on adequate training data, regularization, and model design.

4.2. Discussion

The comparative analysis of the supervised learning algorithms indicates significant information about the behavior and their applicability to high dimensional data classification

ISSN: 3142-788X
PP: 01-14.

processes. Linear models, especially the Logistic Regression, and the Linear Support Vector Machines models show consistently good results mainly because of the way they exploit the use of regularization. Regularization reduces overfitting and improves generalization because it restricts the complexity of the models, which is important in situations where the number of features is much larger than the number of training cases. Such models have advantaged that they have a convex optimization representation that renders stable solutions, and effective training, as they are a dependable option when the dataset is large, and with high dimensionality. The Support Vector Machines particularly the linear ones have a higher propensity to generalize in a sparse and high-dimensional feature space. The margin-maximization principle allows the SVMs to concentrate on the most informative examples during training also referred to as support vectors and thus less sensitivity to noises and irrelevant features. This feature is especially beneficial in applications where feature sparsity is common like text classification and bioinformatics. Upon improved generalization however, there are frequently increased computational needs than linear classifiers, particularly during model training. Similar techniques as Random Forests, which are known as ensemble methods, can provide better robustness by combining the predictions of various base learners. This combination approach has been shown to be effective in decreasing variation and is better at predicting features, particularly when complex interactions between features are at play. In spite of these benefits, ensemble models are computationally expensive, both during training and inference, and this can be a limiting factor in highly dimensional or resource limited settings. And they have more complex models which can diminish interpretability. Neural networks have high representational power, and in competitiveness competitive predictive accuracy, but they are very sensitive to hyperparameter optimization and access to large amounts of labeled data in high-dimensional requirements. Lack of proper regularization and feature preprocessing can lead to overfitting and erratic training of neural networks. It therefore follows that in cases where data is limited to learn, deep models can be very powerful in their learning capabilities, whereas simpler models can offer more dependable and efficient solutions to high-dimensional learning problems.

5. CONCLUSION

This paper has comprised an elaborate comparative analysis of well renowned supervised learning algorithms used to classify high dimensional data. Based on systematically assessed performance metrics at a number of generating features, the study shows a serious impact of feature dimensionality, sample size, and computation limit on the performance of algorithms. The results clearly show that there is no single best learning algorithm that can be used in all high-dimensional situations. Model selection should be instead carefully instructed by the inherent nature of the data, such as sparsity, noise, and the class distribution, and some practical vectors, such as training time, scalability and interpretability needs. The results of the experiment addressed here establish that linear models and margin-based classifiers, including Logistic Regression and linear Support Vector Machines, still remain effective and dependable baselines in high-dimensional learning tasks. This is because they are strong since they have good regularization control systems to limit the complexity of modeling and overfitting, even when the number of features is significantly larger than the number of training examples. These models provide good tradeoff of predictive accuracy,

ISSN: 3142-788X
PP: 01-14.

computational efficiency and interpretability among others, which are quite suitable in real-life scenarios, where transparency and scalability are fundamental. Random Forests are also ensemble methods that achieve better predictive power and robustness because they are trained to learn scores of interactions among features. Nevertheless, this benefit is traded off by a high rate of computational overhead and decreased interpretability that could restrict their use in large scale or time-dependent settings. The approaches based on neural networks have high representational capacity and good classification performance in case of adequate labeled data and computation time. However, sensitivity to hyperparameter configuration, dimensionality of feature space, and data sparsity implies that models should be carefully designed and regularised. In regimes with high dimensionality and small sample sizes, simpler models can be used to outperform deeper ones because the biaserror could be more favorable. The way forward in future studies will be the creation of hybrid learning systems that will be interpretable, efficient like linear models and expressive like non-linear ones. There will be a significant role of selecting the features and dimensionality reduction automated techniques which especially need to be combined into the learning process to enhance scalability and robustness. Also, developing scalable learning algorithms that can operate on an ultra-large-dimensional data, as well as distributed and online learning methods, is a promising research direction in the field of high-dimensional supervised learning studies.

REFERENCES

[1] R. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton, NJ, USA: Princeton Univ. Press, 1961.
[2] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, vol. 1, pp. 1–32, 2000.
[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., New York, NY, USA: Springer, 2009.
[4] J. Friedman, "Regularized discriminant analysis," *J. Amer. Statist. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.
[5] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
[6] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
[7] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed., New York, NY, USA: Springer, 1999.
[8] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[9] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. ECML*, Berlin, Germany: Springer, 1998, pp. 137–142.
[10] A. McCallum and K. Nigam, "A comparison of event models for Naive Bayes text classification," in *Proc. AAAI Workshop on Learning for Text Categorization*, 1998, pp. 41–48.
[11] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge, U.K.: Cambridge Univ. Press, 2003.
[12] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[13] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, 2016.
[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[15] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.