GROUP WORK PROJECT 02
GROUP NUMBER: 7820

MScFE 632: Machine Learning in Finance

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Egor Prokhodtsev	United Kingdom	e.prokhodtsev@gmail.com	X
Regulavalasa Krishna Vamsi	India	krishnavamsi8262@gmail.com	
Enock Joseph Mwesigwa	Uganda	mwesigwaejoseph@gmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an “X” above).

Team member 1	**Enock Joseph Mwesigwa**
Team member 2	**Regulavalasa Krishna Vamsi**
Team member 3	

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed.
Note: You may be required to provide proof of your outreach to non-contributing members upon request.

Egor Prokhodtsev was reassigned to another group on 7th December 2024, and so did not contribute to the GWP.

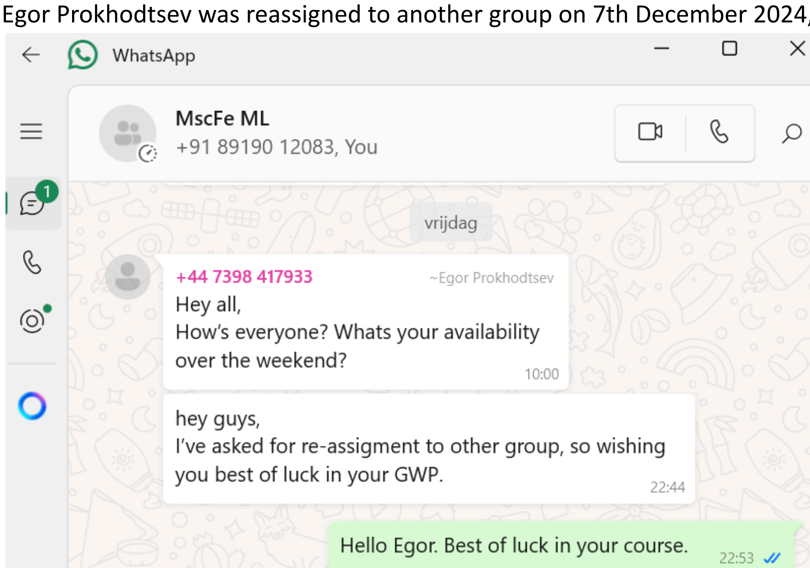


Table of Contents

Step 1

Team Member A - support vector machines (Category 6)

Basics

A Support Vector Machine (SVM) is a supervised learning model used for solving classification and regression problems such as predicting stock returns, bankruptcy, and loan defaults, and is therefore important in risk assessment, and can inform investment strategies. It finds a linear hyperplane that classifies data points in the N-dimensional space where N is the number of features. The SVM focuses on maximising the margin between the data points of different classes to accurately classify the data points with no overlap. As such, hyperparameters can be applied to the hyperplane to adjust it to handle non-linear data.

Keywords:

Hyperplane, Kernel, Margins, Support vectors, Regularisation, Feature space, Classification, Regression

Team Member B - Linear Discriminant Analysis (Category 5)

Basics

Linear Discriminant Analysis is a supervised learning technique, primarily used for classification and dimensionality reduction. It aims to find a linear combination of features that best discriminates between two or more classes of data.

LDA does so by maximizing class separation while reducing the variance inside each class. It finds this projection matrix in order to lower the dimensionality of the given data, from which a decision boundary can then be made that is linear on the new plane for unseen data.

Keywords:
Class separation, Classification, Dimension reduction, Lower-dimension space, Linear decision boundary, Projection matrix.

Step 2

Team Member A -SVM (Category 6)

Advantages

- SVMs are effective in binary classification, making them very applicable to the prediction of bankruptcy and loan default, making them effective risk assessment tools.
- SVMs can handle many asset classes at the same time, enabling the prediction of a desired market condition based on various market conditions.
- When an appropriate regularisation parameter c is used, SVMs are effective in preventing model overfitting in the case of high-dimension data, making them effective in analysing various financial data accurately.
- SVMs can effectively handle non-linear data by use of kernels such as RBF and polynomial and so can used to model non-linear financial relationships on top of also modeling linear data.
- SVM is more robust to outlier influence than linear regression models because it focuses more on the data points close to the margin. This is important in the financial sector where data usually has outliers and extreme events.
- SVMs are efficient in processing high amounts of data quickly with a moderate computational cost, making them effective in algorithmic trading strategies that make rapid trading decisions.

Computation:

In this section, we use SVM to predict the daily stock price movements of Amazon into two classes (that is up (1) or down (0). We used five years of historical data from Yahoo Finance.

Features used include:

- Returns - daily percentage changes in the closing price
- SMA_20 and SMA_50 - Simple moving averages over 20 and 50 days respectively

- RSI - Relative strength index calculated using a 5-day rolling window (assumed a 5-day trading week). This provides information on the price momentum.

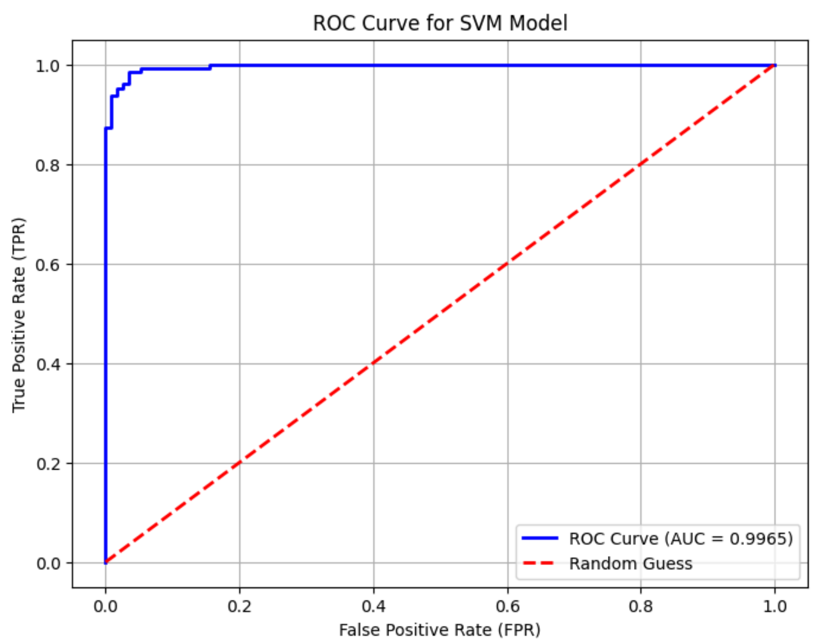
Target is the variable indicating whether the price moved up (1) or down (0) on the following day of trading.

We used the 80:20 split for the data and standardised the features. We applied a radial basis function (RBF) kernel due to non-linear data. We used regularisation parameter c=1 to balance maximisation of the margin with classification accuracy.

Results

Model accuracy is 97%! The classification report is below as well as the ROC curve.

Classes	**Precision**	**recall**	**f1-score**	**support**
0 (Down)	0.98	0.96	0.97	116
1 (Up)	0.96	0.98	0.97	126



All performance indicators show very high performance of the SVM model in predicting Amazon's daily stock price movements with 97% accuracy. The model is good at discriminating between the two classes as shown by the very high AUC value. Far better than the random guesser.

This model is very good at predicting short-term price trends, enabling investors to make quick buy or sell decisions.

Disadvantages

- SVMs are computationally expensive when training them with large datasets. This limits their use to reasonably financially large firms and organisations.
- SVM performance depends on the choice of the kernel and its parameters. An inappropriate kernel, such as a linear kernel for polynomial data can lead to poor model performance.
- SVM is highly sensitive to its hyperparameters such as the regularization parameter C. Accurate tuning is always required to be able to prevent model overfitting and underfitting.
- SVM can be considered a black box model due to the difficulty in interpreting its mode of action, especially when non-linear kernels (such as RBF) are used. This contrasts with the linear regression model which is easier to interpret and explain.
- SVM is sensitive to the scaling of the features where normalization or standardisation needs to be done before the model can analyse the data. The model will perform poorly if features are not scaled properly.

Equations

1. **Maximising the margin**

SVM seeks to maximize the margin, i.e. the distance between the hyperplane and the nearest data point of each class.

$Margin = \frac{2}{||w||}$ where the hyperplane is defined as $w^T x + b = 0$

Where w is the weight vector (coefficients of the hyperplane), $||w||$ is the Euclidean norm of the weight vector, and b is the bias term.

2. **Optimization**

SVM uses convex optimization:

$min\ of\ w, b \frac{1}{2}||w||^2$ subject to $y_i(w^T x_i + b) \geq 1, \forall i$

Where x_i is the input feature vector of the i-th sample, and y_i is the label of the i-th sample(+1 or -1) to ensure a distance of at least 1 between the hyperplane and the nearest data points of each class.

3. **Soft margin for non-separable data**

We add a slack variable for non-linearly separable data to allow a soft margin that allows some misclassifications. This transforms the optimization problem into:

$min\ of\ w, b, \xi \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i$ subject to $y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i$

Where ξ_i is the slack variable for the i-th sample, and C is the regularization parameter that balances the margin and the miscalculation penalty. If C is too high, the margin is minimized, and vice versa.

4. **Kernel functions**

SVM uses kernels to map input of non-linearly separable data into higher dimensional space where it is linearly separable. The decision function is expressed as:

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b)$$

Where:

- α_i are Lagrange multipliers
- $K(x, x_i)$ is the kernel function
- Linear kernel: $K(x, x') = x^T x'$
- Polynomial kernel: $K(x, x') = (x^T x' + c)^d$
- RBF kernel: $K(x, x') = exp(-\gamma||x - x'||^2)$ where γ is the standard deviation of the Gaussian distribution

5. **Probability estimation in SVM**

Probability in the computation was set to True and so SVM used Platt scaling to convert the decision function $f(x)$ into probabilities for purposes of calculating ROC.

$P(y = 1|x) = \frac{1}{1+exp(Af(x)+B)}$ where A and B are obtained through cross-validation.

6. **Decision boundary**

The decision boundary is $f(x) = 0$ and so becomes:

$$\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b = 0$$

Features

- SVMs find the optimal hyperplane that maximises the margin between classes enabling accurate classifications.
- SVMs use support vectors which are a subset of the training data points to make decisions, making them efficient in terms of memory when handling large data.

- SVMs are kernel flexible, in that they can be applied with any of the many kernels such as RBF, polynomial, and linear, making SVMs highly adaptable to different data patterns.
- The regularisation parameter C enables modification of the balance between margin maximization and misclassification, enabling prevention of the model overfitting volatile financial data.
- SVMs are effective in handling high-dimensional data with many features, making them a good choice for financial problems that involve multiple features.
- SVMs are applicable to both binary and multi-class classification and hence can be used for both single stock management as well as portfolio management.
- SVMs are non-parametric in nature, and so can be used with datasets even where the distribution is not known, as is the case with most financial data.
- SVMs can handle imbalanced data through class weighting that can adjust the decision function. This extends their applicability to diverse datasets.

Guide: Inputs and outputs

Inputs

1. Data consisting of features (X) and target (y)
2. Kernel type such as linear, polynomial, RBF, sigmoid
3. Regularization parameter C. High C minimises misclassifications and vice versa.
4. Kernel parameters where needed. For the polynomial kernel, Degree d and coefficient c are needed. RBF kernel requires the gamma parameter γ
5. Scaled or normalised features using scalers such as standard, min-max
6. Hyperparameter values for tuning the parameters C, γ and any other kernel parameters

Outputs

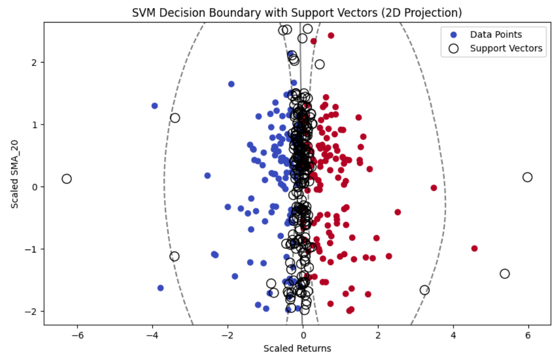
1. Trained model
2. Support vectors
3. Dual coefficients α_i for support vectors
4. Intercept (bias b)
5. Decision function
6. Predicted class labels

Hyperparameters

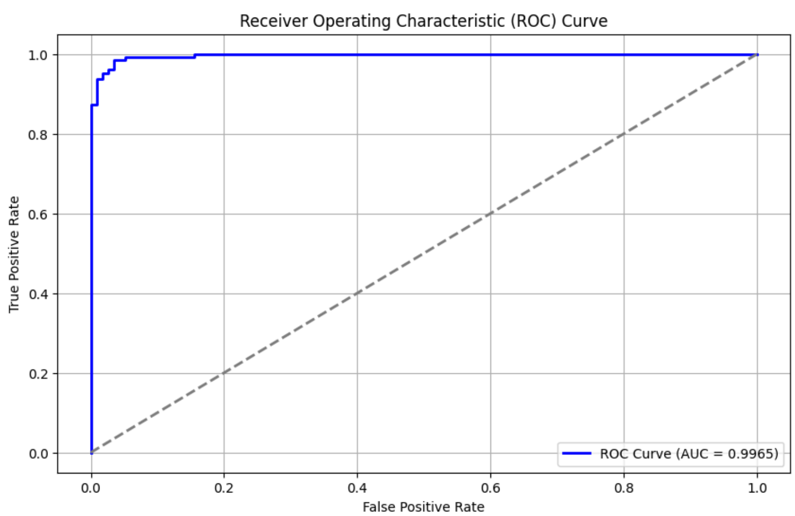
1. Regularisation parameter C
2. Kernel type
3. Kernel-specific parameters such as degree d and coefficient c for the polynomial kernel, and gamma γ for RBF kernel
4. Class weights

Illustration

The first illustration shows the decision boundary that separates classes in the feature space. The centre line is the hyperplane, with the red and blue data points representing the two target classes.



The second figure displays the ROC curve that illustrates the performance of the particular SVM model we used in the computation. As can be observed, the model was highly accurate at predicting whether the stock price of Amazon would increase or decrease in the following trading day.



Journal

The journal article by (Kurani, Doshi and Vakharia) describes the application of SVM in predicting the stock prices of Coca Cola company. It performed better than the linear regression model that was applied to the same task. This emphasizes its applicability to non-linear financial data.

- Kurani, Akshit, et al. "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) Network (ANN) and Support Vector

Machines (SVM)." Annals of Data Science 10 (2023): 183-208. <https://doi.org/10.1007/s40745-021-00344-x>.

Team Member B - Linear Discriminant Analysis (Category 5)

Advantages of Linear Discriminant Analysis

1. Simplicity: LDA is a simple and intuitive algorithm.
2. Computational Efficiency: It is computationally efficient, and thus well suited for large data.
3. Class Separability: After projecting the data on a lower dimension, LDA maintains class separability, enhancing model generalization.
4. Handling High-Dimensional Data: LDA can effectively handle high-dimensional data, which makes it applicable to many real-world applications

Computation:
Predictive model of a Linear Discriminant Analysis (LDA) of NVDA Stock Returns. For this model, the input feature contained stock returns on Apple, Microsoft, IBM, Cisco, and Google. My goal was classifying NVDA returns as a 1 in case they crossed 2%, and 0 otherwise.

Additional information on implementation and results can be found in the respective Jupyter notebook.

Disadvantages

1. Normality Assumption: LDA assumes that the data in each class comes from a normal distribution, an assumption that can drastically affect the results if it turns out to be violated.
2. Assumption of equal covariance: All classes are assumed to have the same covariance matrix under LDA, which may be misleading if all classes have varying variances.
3. The Presence of Outliers can distort means and variances that are necessary for defining boundary conditions, often causing incorrect classes.
4. It is not always easy to interpret the linear combinations of features that emerge from LDA, especially in high dimensions.

Equations:

- Standardization of Features: The dataset is standardized by making use of the StandardScaler. This transforms the data such that each feature will have a mean of 0 and a standard deviation of 1.To the best of LDA, its objective is the determination of that linear combination of features which classifies the different classes with optimum separation by the maximization of the ratio between-class variance within-class variance in order to result in the appropriate projection of the data for proper classification.

respectively.

$$\mu'_1 = \frac{1}{n}\Sigma v^T x_i$$

$$= v^T \mu_1$$

Similarly;

$$\mu'_2 = v^T \mu_2$$

Where:

x_i ,i =1,2,3…n represent the data points

v^T is the set of unit vectors that discriminate between the two classes.

To find the unit vector v that maximizes the discrimination between the classes,

We will solve Problem:

$$max\ (v: ||v|| = 1)\ |\ \mu'_1 - \mu'_2\ |$$

The sample class c1 and c2 have the following variances:

$$s'_1 = \sum (y_i - \mu_1)^2$$

Where $y_i \in c_1$

$$s'_2 = \sum (y_i - \mu_2)^2$$

Where $y_i \in c_2$

Where : $y_i = v^T x_i$

The optimal choice of v should be such that:

- $(\mu'_1 - \mu'_2)^2$ is large and
- $(S'_1)^2$ $(S'_2)^2$ are both small

To achieve this, we solve the problem:

$$max\ (v: ||v|| = 1)\ \frac{(\mu'_1 - \mu'_2)\ 2}{(s'_1)\ 2 + (s'_2)\ 2}$$

Features:

1. Sensitivity to Outliers: LDA might be sensitive to outliers, hence the performance can degrade and classify inaccurately in certain cases.
2. Normality Assumption: LDA assumes that the data in each class is normally distributed.
3. Equal Covariance Assumption LDA assumes equal covariance matrices across all classes
4. Dimensional Reduction: In LDA, the original space is projected in a lower space for improving separation between classes.
5. It provides a linear combination of features best differing the classes being separated.

Guide

Inputs:

- Feature Data: input variables used for training the model.
- Class Labels: target labels attached to each training sample.

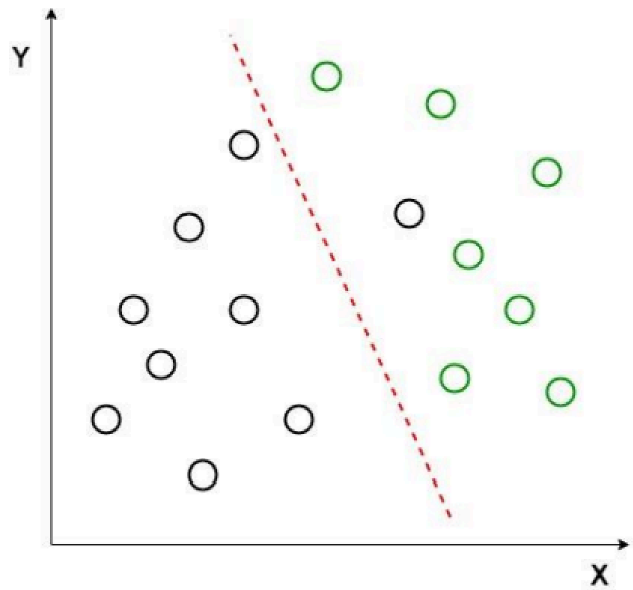
Outputs

- Predicted Labels: class prediction of new data samples
- Discriminant Values: A score which states how much a data point supports its class label.
- Feature Weights: the coefficients attached to each feature defining the best linear combination to be used in the separation of classes.

Hyperparameters

- LDA is a simple algorithm that usually does not require fine-tuning of hyperparameters. Yet, the use of model regularization can be improved through the following:
    - Incorporation of Prior Probabilities: If there is prior knowledge regarding the likelihood of a point in the data falling under a particular class, such probabilities could be used to adjust the model.
    - Standardizing Features: When features scale differently, standardizing helps to ensure each feature contributes equally to the model and thus improves its performance.

Illustrations:



The above illustration displays how LDA uses both X and Y axes to create a new axis that maximizes the separation of the data into two classes and hence, reduces the 2D graph into a 1D graph.

Journal:

○ Tharwat, Alaa et al, May 2017 “Linear discriminant analysis: A detailed tutorial”, ResearchGate. DOI: [10.3233/AIC-170729](https://doi.org/10.3233/AIC-170729)

Step 3. Technical section

Support Vector Machines (SVM)

Hyperparameters are those whose values are not derived or learned from the data but are set by the researcher before the model is trained. They are used in the model and influence the model performance. Because of this, tuning them to appropriate and optimum values is very important in ensuring the model performs well.

In SVM, we have several hyperparameters including the kernel type, the regularisation parameter C, and gamma. We can use several methods for tuning them.

- Grid search: Here, a hyperparameter space is provided to the algorithm that manually iterates over all combinations and chooses the optimum combination based on model performance characteristics such as mean absolute error. Grid search is ideal with a small hyperparameter space as computational time increases exponentially with the increase of the hyperparameter space.

- Random search: This is an improvement on grid search in that the combinations that are analysed are randomly chosen, hence limiting the number of samples. This allows for bigger hyperparameter spaces to be used without extreme increase in computational time.
- Bayesian optimisation: This uses a probabilistic model of the objective function to efficiently identify optimal hyperparameters. It is expensive due to the additional model cost, but it is most efficient for use in very large hyperparameter spaces.
- Cross-validation: It is used to ensure unbiased hyperparameter tuning by validating the tuned parameters with out-of-sample data.

An illustration of the application of grid search to tuning hyperparameters for the SVM is in the attached Python notebook.

Challenges in tuning SVM hyperparameters

- Tuning attracts a high computational cost, with the cost increasing exponentially with an increase in the size of the dataset, especially when tuning non-linear kernels such as RBF.
- Kernel selection requires good domain knowledge to know which kernel applies to the data set. This constrains the model performance to the practitioner's expertise and affects model predictions.

Linear Discriminant Analysis (LDA):

The LDA algorithm operates without the need for hyperparameters, reflecting its straightforward and linear design.

Step 4. Marketing Alpha

Support Vector Machines

Support Vector machines (SVM) can be applied in finance as machine learning models that excel at modelling complex patterns in financial data while also remaining robust to noise.
SVMs are less sensitive to noise in the financial data which enables them to accurately and reliably model volatile assets such as Amazon's stock. SVMs, with their arsenal of kernels such as RBF, polynomial, and sigmoid, can model non-linear relationships in financial data, such as

dependencies between technical indicators and future returns. SVMs are very effective in handling high-dimensional financial data, ensuring accuracy and interpretability. SVMs are also flexible in that they can perform regression (such as forecasting returns), and also classification (such as predicting up or down price movements). This makes them applicable to many financial modeling problems.

We used SVM to predict the daily price movements of Amazon stock and it achieved a 97% accuracy in its predictions showing its high predictive power. The AUC of 0.9965 indicates a very high classification capability which translates to accurate and highly reliable predictions.

In financial trading, SVM can provide actionable information on market data enabling profitable allocation of investment resources. Since SVMs are accurate in modeling non-linear patterns, they can effectively identify patterns preceding market downturns, hence informing risk mitigation strategies. Their scalability to different asset classes and trading frequencies also makes them very essential in modern financial trading strategies.

Linear Discriminant Analysis (LDA):

LDA can be viewed as an alternative to LDA and used for classification and not for regression.

It reduces the dimensionality of data by projecting it onto a lower-dimensional space without losing significant data.

The financial time series data are generally non-normal, which affects the performance of LDA. Applying the model to predict the returns of NVDA stock yields an accuracy of about 58%, based on LDA's normality assumption and the inherent non-linearity of stock returns.

Step 5. Learn more

Support Vector Machines

These journal articles show the application of SVMs in finance, emphasizing their strengths and features, highlighting their weaknesses, and how they can be improved.

1. Kurani, Akshit, et al. "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) Network (ANN) and Support Vector Machines (SVM)." Annals of Data Science 10 (2023): 183-208. <https://doi.org/10.1007/s40745-021-00344-x>.

2. Liu, Ying and Lihua Huang. "Supply chain finance credit risk assessment using support vector machine–based ensemble improved with noise elimination." International

Journal of Distributed Sensor Networks 16.1 (2020). <https://doi.org/10.1177/1550147720903631>.

3. Teles, Germanno, et al. «Comparative study of support vector machines and random forests machine learning algorithms on credit operation.» *Software: Practice and Experience* 51.12 (2020): 2492-2500. <https://doi-org.kuleuven.e-bronnen.be/10.1002/spe.2842>.

4. Wang, Tianhui, Renjing Liu and Guohua Qi. "Multi-classification assessment of bank personal credit risk based on multi-source information fusion." Expert Systems with Applications 191 (2022). <https://doi.org/10.1016/j.eswa.2021.116236>.

5. Zhang, Hao, et al. "A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance." Research in International Business and Finance 58 (2021). <https://doi.org/10.1016/j.ribaf.2021.101482>.

LDA:

Dr. Guangliang Chen, “Math 253: Mathematical Methods for Data Visualization Linear Discriminant Analysis (LDA)”, San José State University.
https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec11lda.pdf

Linear Discriminant Analysis in Machine Learning
https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/

James, Gareth, et al. An Introduction to Statistical Learning: With Applications in R. 2nd ed, Springer, 2021. Chapter 4.4.

Step 6. Comparing models

Features	Linear Discriminant analysis	support vector machines
Handles Missing data	Handles Poorly	Handles Poorly
Scalability	Scales well	Scale Moderately
Overfitting	Prone to Overfitting	Less Prone to Overfitting
Outlier sensitivity	Sensitive to outliers	Robust to outliers
Computational complexity	Highly interpretable	can be computationally expensive
Non-linear relationships	Very Limited, works best with linear data	Very Effective by using kernel functions
Feature importance	Provide clear feature importance	Limited feature importance
Multiclass Classification	Natural Support for Multiclass classification	Requires 1 v 1 or 1 v multi-approach
Regularization	Limited, uses covariance matrix instead	Built-in regularization

References

1. Kurani, Akshit, et al. "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) Network (ANN) and Support Vector Machines (SVM)." Annals of Data Science 10 (2023): 183-208. <https://doi.org/10.1007/s40745-021-00344-x>.
2. Liu, Ying and Lihua Huang. "Supply chain finance credit risk assessment using support vector machine–based ensemble improved with noise elimination." International Journal of Distributed Sensor Networks 16.1 (2020). <https://doi.org/10.1177/1550147720903631>.
3. Teles, Germanno, et al. «Comparative study of support vector machines and random forests machine learning algorithms on credit operation.» *Software: Practice and*

Experience 51.12 (2020): 2492-2500. <https://doi-org.kuleuven.e-bronnen.be/10.1002/spe.2842>.

4. Wang, Tianhui, Renjing Liu and Guohua Qi. "Multi-classification assessment of bank personal credit risk based on multi-source information fusion." Expert Systems with Applications 191 (2022). <https://doi.org/10.1016/j.eswa.2021.116236>.
5. Zhang, Hao, et al. "A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance." Research in International Business and Finance 58 (2021). <https://doi.org/10.1016/j.ribaf.2021.101482>.