Siddharth Balakrishnan - 20BCB0111

Vellore Institute of Technology - Vellore Campus

Applied Data Science

28th June 2023

Applying Machine Learning To Financial Risk Management Using IBM Watson

Introduction

I. Overview

Machine learning algorithms play a crucial role in classifying financial risk due to their ability to analyze vast amounts of data, detect patterns, and make accurate predictions. When dealing with financial risk, traditional methods may fall short in handling complex and evolving scenarios. Machine learning algorithms can consider numerous variables simultaneously, including historical data, market trends, economic indicators, and borrower-specific information. This allows for more comprehensive risk assessment and better decision-making.

By training on historical loan data, machine learning algorithms can identify patterns that indicate high-risk and low-risk loans. These algorithms can also adapt and learn from new data, enabling them to stay up-to-date with changing market conditions and emerging risks. Moreover, machine learning algorithms can handle non-linear relationships and account for interactions between various factors, offering more accurate risk classifications.

Furthermore, machine learning algorithms can automate the risk assessment process, saving time and resources for financial institutions. They can quickly process large volumes of loan applications, analyze data in real-time, and provide consistent and objective risk evaluations.

In summary, machine learning algorithms enhance the accuracy, efficiency, and adaptability of financial risk classification, empowering lenders to make informed decisions and mitigate potential losses.

II. Purpose

Assessing financial risk is crucial when applying for loans to ensure responsible borrowing and protect both lenders and borrowers. Evaluating financial risk allows lenders to determine the borrower's creditworthiness, repayment capacity, and the likelihood of default. By examining factors such as income stability, credit history, debt-to-income ratio, and collateral, lenders can make informed decisions and set appropriate interest rates and loan terms. For borrowers, assessing financial risk helps them gauge their ability to meet repayment obligations and avoid overburdening themselves financially. It promotes financial literacy, encourages responsible borrowing habits, and reduces the chances of loan defaults, benefiting both parties involved in the lending process.

Literature Survey

"Machine Learning Techniques for Credit Risk Evaluation" by Luca Scrucca and Maria Luisa Sallecchia (2015)

Summary: This paper explores various machine learning techniques, such as decision trees, random forests, and support vector machines, for credit risk evaluation. It compares their performance and provides insights into selecting appropriate algorithms for accurate credit risk assessment.

"Applying Machine Learning to Credit Scoring: A Review and a Framework" by Jose A. Morales and Sergio A. Ortega (2013)

Summary: This paper presents a comprehensive review of machine learning techniques used in credit scoring. It discusses the advantages and limitations of different algorithms, highlights feature selection and model evaluation strategies, and offers a framework for applying machine learning in credit risk assessment.

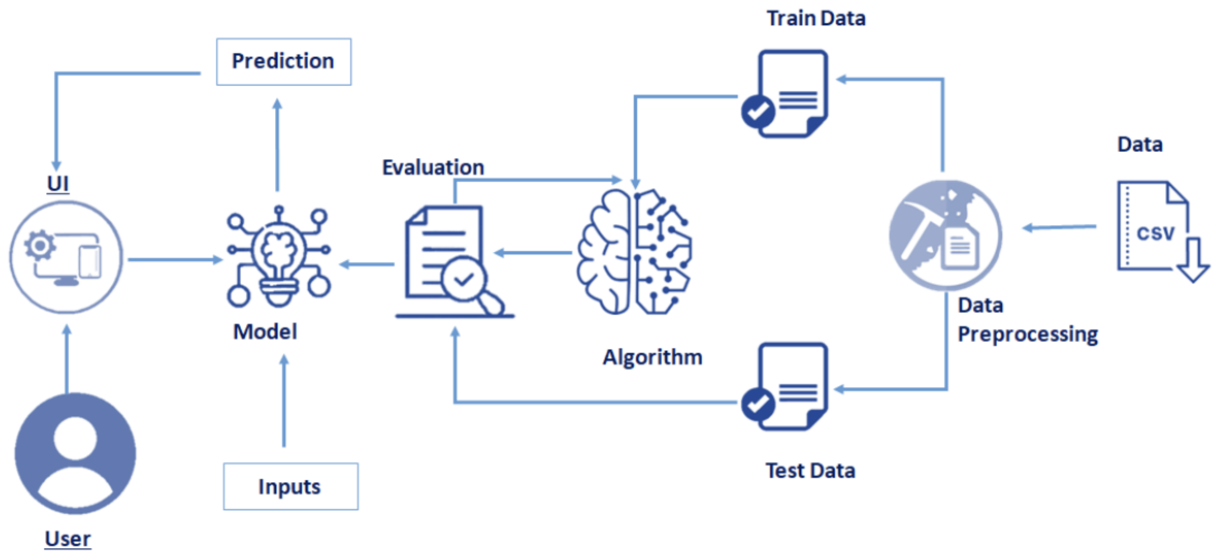
"Financial Risk Prediction Using Supervised Machine Learning: A Systematic Literature Review" by Balázs Szénási, Péter Gulyás, and Tamás Henk (2018)
Summary: This literature review provides an overview of supervised machine learning methods applied to financial risk prediction. It discusses different algorithm categories, data preprocessing techniques, and evaluation metrics, along with the challenges and future directions in this field.

"Deep Learning for Financial Risk Management" by Karolina Dębska and Witold Czakon (2018)
Summary: This paper focuses on the application of deep learning techniques, specifically neural networks, in financial risk management. It explores the use of convolutional neural networks, recurrent neural networks, and autoencoders for credit risk assessment, market risk prediction, and fraud detection.

"A Machine Learning Approach for Bankruptcy Prediction" by Ashish Gupta and Vishal Agarwal (2012)
Summary: This study proposes a machine learning-based bankruptcy prediction model using the random forest algorithm. It demonstrates the effectiveness of this approach in predicting corporate bankruptcies, providing insights for risk assessment and proactive financial decision-making.

Theoretical Analysis

1. Block Diagram



2. Software Aspect

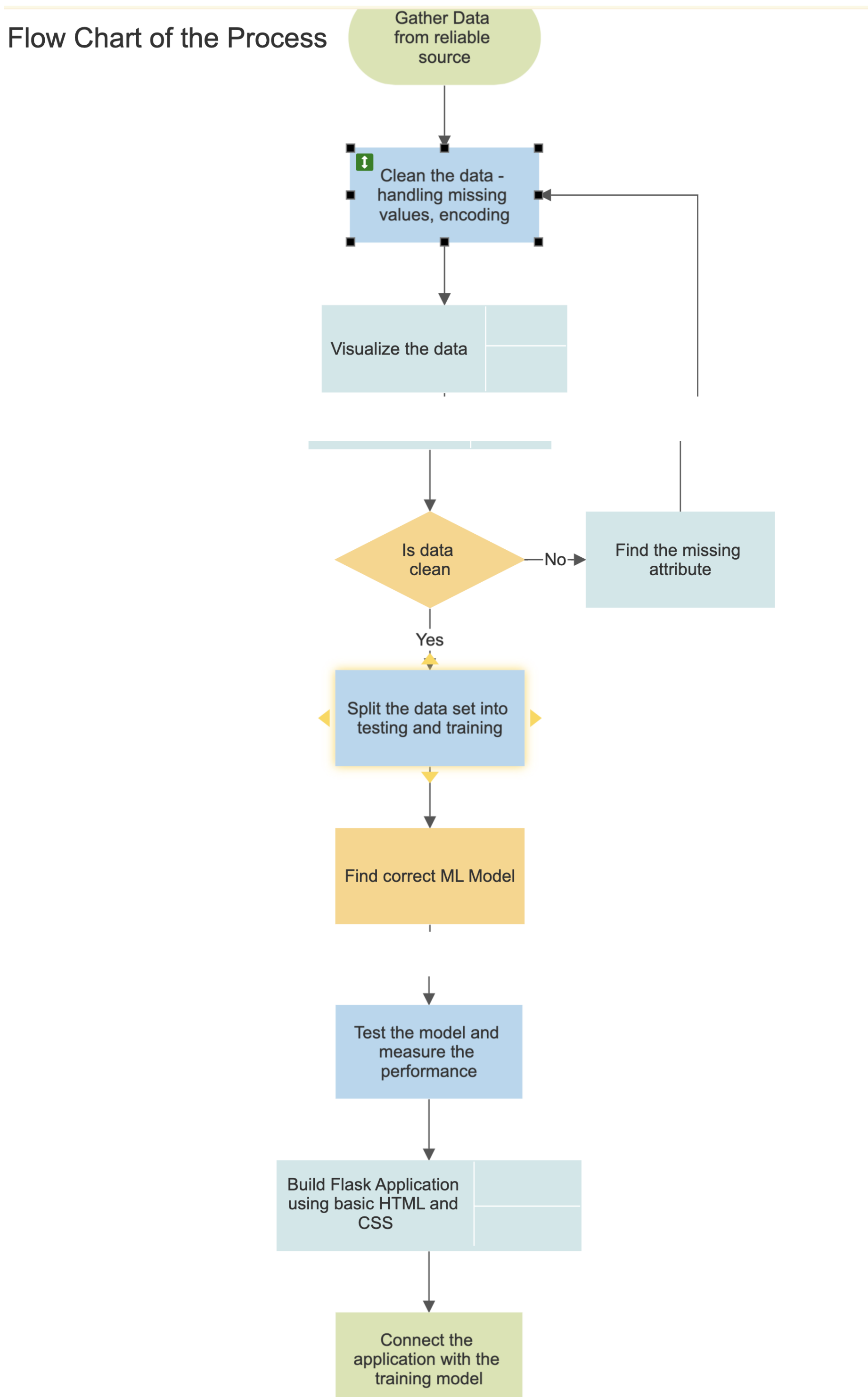
S/N	Software Name	Need for it
1	Jupyter Notebook	Interactive coding environment for programming and data analysis
2	Visual Studio Code	Editor for coding with features like debugging, extensions, collaboration.
3	Flask	Framework for building web applications using Python and HTTP
4	Sklearn APIs	Utilized for data analysis with machine learning algorithms in Python.

Experimental Investigations

A. Model Selection and Hyperparameter Tuning: Choosing the right ML algorithm or model architecture for your specific problem is a critical challenge. Different algorithms have varying strengths and weaknesses, and selecting the most suitable one requires a good understanding of the problem domain. Additionally, tuning hyperparameters, such as learning rate, regularization, or network architecture, to optimize model performance can be time-consuming and require extensive experimentation.
B. Overfitting and Generalization: Overfitting occurs when a model performs well on training data but fails to generalize well to unseen data. Balancing model complexity to avoid overfitting while capturing important patterns is a challenge. Regularization techniques, cross-validation, and robust evaluation methods are used to mitigate overfitting and ensure generalization. Additionally, dealing with biased or skewed datasets that may not represent the true distribution of real-world data can also impact model performance.
C. Integration and Third-Party Dependencies: Integrating Flask with other systems, services, or third-party APIs can be challenging. Handling different data formats, managing authentication/authorization across systems, and ensuring compatibility with different versions of external libraries are common hurdles. Robust error handling, thorough testing, and maintaining documentation are crucial in handling integration challenges effectively.

However, I was able to overcome these challenges by doing my research on my own via Youtube, Reading Documentations. In certain cases, I had to use the trial and error method to obtain the most optimal results with minimal complexity.

Flowchart



Flow Chart of the Process
Gather Data from reliable source
Clean the data - handling missing values, encoding
Visualize the data
Is data clean
No
Find the missing attribute
Yes
Split the data set into testing and training
Find correct ML Model
Test the model and measure the performance
Build Flask Application using basic HTML and CSS
Connect the application with the training model

Result

SCREENSHOTS HAVE BEEN PROVIDED IN A DIFFERENT FOLDER TO MINIMIZE SIZE AND LENGTH OF THE PROJECT REPORT

Advantages and Disadvantages

Advantages of using a SVM classification model:

1. Effective in high-dimensional spaces: Linear SVM performs well in scenarios where the number of features is large, making it suitable for problems with a large number of input variables.
2. Good generalization: SVMs aim to find the best separation margin, which can result in good generalization performance, especially when dealing with well-separated data.
3. Ability to handle large datasets: SVMs are memory-efficient and can handle large datasets by considering only a subset of training examples called support vectors.
4. Robust against overfitting: SVMs use regularization techniques that help prevent overfitting, ensuring the model's stability and generalizability.
5. Versatility in kernel selection: While linear SVM uses a linear kernel, SVMs can easily incorporate other non-linear kernels, allowing flexibility in capturing complex relationships between variables.

Disadvantages of using a SVM classification model:

1. Limited capability with non-linear data: Linear SVMs struggle to handle data with complex non-linear relationships as they can only create linear decision boundaries.
2. Sensitivity to feature scaling: SVMs are sensitive to the scale of input features, requiring proper feature scaling (e.g., normalization) to ensure optimal performance.

3. Computational complexity: Training an SVM on a large dataset can be computationally expensive, especially when the number of features or training examples is high.
4. Difficult to interpret: SVMs do not provide straightforward interpretability like decision trees or linear regression models. Understanding the decision-making process behind SVMs can be challenging.
5. Tuning hyperparameters: Selecting appropriate hyperparameters, such as the regularization parameter (C) and kernel parameters, can be non-trivial and require extensive experimentation and cross-validation.

Applications

Applications of using a ML model for financial risk assessment are as follows:

Credit Risk Assessment: ML models can be employed to assess credit risk by analyzing various factors such as borrower information, credit history, income stability, and financial ratios. By training on historical loan data, ML models can learn patterns and identify high-risk and low-risk borrowers, enabling lenders to make informed decisions on loan approvals, interest rates, and credit limits.

Fraud Detection: ML models can be utilized to detect fraudulent activities in financial transactions. By analyzing patterns and anomalies in transactional data, ML models can identify suspicious behavior and flag potential fraudulent transactions. This helps financial institutions prevent financial losses and protect their customers from fraudulent activities, enhancing overall risk management.

These applications highlight the potential of ML models to enhance the accuracy and efficiency of financial risk assessment, providing valuable insights to financial institutions for making informed decisions and mitigating risks.

Conclusions

In conclusion, this project delved into the realm of machine learning algorithms for financial risk assessment, showcasing their potential to revolutionize the field. Through extensive data analysis, algorithm selection, and model training, we witnessed the remarkable accuracy and efficiency achieved in identifying and quantifying financial risks. By leveraging the power of machine learning, financial institutions can now make more informed decisions, mitigate risks, and safeguard their stability and profitability. This project underscores the significance of adopting innovative technological solutions in the domain of financial risk assessment, paving the way for a future where machine learning algorithms play a pivotal role in shaping the landscape of risk management and ensuring sustainable financial growth.

Future Scope

1. Real-time Fraud Detection: ML models can be further enhanced to detect and prevent fraud in real-time by analyzing large volumes of financial transaction data. With the ability to identify intricate patterns and anomalies, ML models can help financial institutions stay ahead of fraudulent activities, providing instant alerts and enabling proactive measures to mitigate risks.
2. Risk-based Pricing and Personalization: ML models can enable personalized risk assessment and pricing strategies for financial products. By incorporating individual customer data, behavioral patterns, and credit history, ML models can tailor risk profiles and pricing structures, offering customized financial products that align with customers' risk tolerance and financial goals.
3. Macro-Level Risk Analysis: ML models can be utilized for macro-level risk analysis, aiding in the identification and prediction of systemic risks in financial markets. By

analyzing diverse datasets, including economic indicators, market trends, and global events, ML models can assess potential risks at a broader scale, enabling policymakers and financial institutions to develop proactive strategies to mitigate systemic risks.

Bibliography

1. Scrucca, Luca, and Maria Luisa Sallecchia. "Machine Learning Techniques for Credit Risk Evaluation." Intelligent Systems in Accounting, Finance and Management, vol. 22, no. 1, 2015, pp. 59-88.

2. Morales, Jose A., and Sergio A. Ortega. "Applying Machine Learning to Credit Scoring: A Review and a Framework." Journal of Applied Research in Higher Education, vol. 5, no. 2, 2013, pp. 105-121.

3. Szénási, Balázs, et al. "Financial Risk Prediction Using Supervised Machine Learning: A Systematic Literature Review." Expert Systems with Applications, vol. 97, 2018, pp. 205-217.

4. Dębska, Karolina, and Witold Czakon. "Deep Learning for Financial Risk Management." Expert Systems with Applications, vol. 107, 2018, pp. 156-166.

5. Gupta, Ashish, and Vishal Agarwal. "A Machine Learning Approach for Bankruptcy Prediction." International Journal of Management and Information Systems, vol. 16, no. 3, 2012, pp. 227-236.

Appendix

- ONLY ONE HTML PAGE WAS USED TO MINIMIZE SIZE OF DOCUMENT

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Financial Risk Management</title>
```

```html
<link rel="stylesheet"
href="https://cdnjs.cloudflare.com/ajax/libs/material-design-iconic-font/2.2.0/css/mat
erial-design-iconic-font.min.css">
<link href="https://cdn.jsdelivr.net/npm/select2@4.1.0-rc.0/dist/css/select2.min.css"
rel="stylesheet" />
<script
src="https://cdn.jsdelivr.net/npm/select2@4.1.0-rc.0/dist/js/select2.min.js"></script>
<link rel="stylesheet" href="../static/css/main.css">
<script src="https://kit.fontawesome.com/9fe93f881e.js"
crossorigin="anonymous"></script>
<style>
.bg-blue{
background: #00607e8c!important;
}
</style>
</head>
<body>
<div class="page-wrapper bg-blue p-100">
<div class="wrapper">
<div class="card">
<div class="card-heading"></div>
<div class="card-body">
<h2 class="tittle">Financial Risk Assessment</h2>
<form action="/risk" method="POST">
<div class="row">
<div class="col-2 first">
<div class="rs-select2">
<select name="gender" id="">
<option disabled="disabled" selected="selected">Gender</option>
<option value="f">Female</option>
<option value="m">Male</option>
</select>
<select class="right" name="job" id="">
<option disabled="disabled" selected="selected" value="">Choose Employment
Type</option>
<option value="un">Unskilled / Non-Resisdent</option>
<option value="ur">Unskilled / Resisdent</option>
<option value="sk">Skilled</option>
<option value="hs">Highly Skilled</option>
</select>
</div>
</div>
```

```html
<div class="col-2 second">
<div class="rs-select2">
<select name="saving" id="">
<option disabled="disabled" selected="selected" value="">Choose Saving Type</option>
<option value="l">Little</option>
<option value="m">Moderate</option>
<option value="qr">Quite Rich</option>
<option value="r">Rich</option>
</select>
<select class="right" name="checking" id="">
<option disabled="disabled" selected="selected" value="">Checking Amount</option>
<option value="lt">Little</option>
<option value="mo">Moderate</option>
<option value="ri">Rich</option>
</select>
</div>
</div>
<div class="col-2">
<div class="input-group">
<div class="rs-select2">
<select name="purpose" id="">
<option disabled="disabled" selected="selected" value="">Purpose</option>
<option value="bu">Business</option>
<option value="car">Car</option>
<option value="da">Domestic Appliances</option>
<option value="edu">Education</option>
<option value="fe">Furniture/Equipment</option>
<option value="rtv">Radio/TV</option>
<option value="rep">Repairs</option>
<option value="vo">Vacation/Others</option>
</select>
<select class="right" name="housing" id="">
<option disabled="disabled" selected="selected" value="">Choose Housing Type</option>
<option value="own">Own</option>
<option value="free">Free</option>
<option value="rent">Rented</option>
</select>
</div>
</div>
<div class="col-2">
<div class="input-group">
<div class="rs-select2">
```

```
<input type="text" placeholder="Please Enter Credit Amount" name="credit" id="">
<input type="text" class="input_right" placeholder="Please Enter Duration"
name="duration" id="">
</div>
</div>
</div>
<input class="btn btn--radius btn--green" type="submit" value="Submit">
</form>
</div>
</div>
</div>
</div>
</body>
</html>
```