Project F9: Estonian Company Analysis

Team Members:

Siim Turban

Kevin Alak

Repository link: https://github.com/Auraplus5/DataSciProject (Public)

Task 2. Business understanding

Identifying Business Goals

Background:

Our project, "Estonian Company Analysis," intends to undertake a comprehensive analysis of Estonian companies, with a focus on sectors such as technology, manufacturing, and services. The purpose is to examine factors like earnings, market dynamics, and regulatory compliance. By leveraging advanced data analytics and machine learning, we aim to uncover trends and patterns that define business success in Estonia.

Business Goals:

Develop a robust machine learning model to analyse the factors influencing Estonian companies' success, considering aspects such as financial performance, market positioning, and regulatory adherence.

Employ predictive analytics to assess the potential success of new businesses in Estonia, considering current market trends and historical business data.

Business Success Criteria:

Creation of a machine learning model that accurately interprets and predicts business performance.

Development of actionable insights that can guide new and existing businesses in strategic decision-making.

Assessing the Situation

Inventory of Resources:

Comprehensive datasets encompassing years of financial records, market data, and regulatory information of Estonian businesses.

Requirements, Assumptions, and Constraints:

Access to up-to-date and comprehensive data from Estonian business registries and market research firms.

Assumption: The economic and regulatory environment will remain stable over the analysis period.

Constraints include limited time for project completion and finite computational resources for extensive data processing and model training.

Risks and Contingencies:

Risks related to data integrity, such as incomplete or outdated information.

Market volatility that could render predictive models less accurate over time.

Contingency plans include regular data validation checks and model updates to adapt to changing market conditions.

Terminology:

"Business success" is quantified in terms of revenue growth, market share expansion, and adherence to regulatory standards.

"Market concentration" refers to the competitive landscape and market share distribution among companies in various sectors.

Costs and Benefits:

Costs involve human resource allocation for research and analysis, and technological expenses for data processing and model development.

The benefits are multi-faceted, including improved understanding of market dynamics, enhanced predictive capabilities for business performance, and valuable insights for stakeholders in the Estonian business community.

Defining Data-Mining Goals

Data-Mining Goals:

To dissect and understand the key determinants of business success in the Estonian market.

To build predictive models that can reliably forecast business outcomes based on market and internal company data.

Data-Mining Success Criteria:

Achieving a high degree of accuracy and reliability in the predictive models.

The ability of the model's insights to be translated into effective business strategies and policies.

Task 3: Data Understanding

Gathering Data

Outline Data Requirements:

Detailed financial records for Estonian companies, including annual revenue, expenses, profit margins, and growth rates, spanning at least the past ten years.

Regulatory compliance records, including histories of compliance or violations, adherence to national and international regulations, and any legal actions taken against companies.

Verify Data Availability:

Engage with Estonian business registries, financial institutions, and market research agencies to secure access to the required financial and market data.

Find the trustable sources to obtain compliance records and relevant legal documents.

Define Selection Criteria:

Focus on data that covers a wide range of industries, from emerging tech startups to established manufacturing firms, ensuring a holistic view of the business environment.

Select data primarily from verified, credible sources to ensure accuracy and reliability.

Ensure that the data is current, ideally not older than a year, to maintain relevance in fast-changing market conditions.

Describing Data

Quantitative Financial Data: In-depth analysis of financial metrics like cash flow statements, balance sheets, and profit and loss statements, providing insights into the financial health and stability of companies.

Qualitative Market Data: Narrative descriptions of market trends, incorporating expert analyses, sector growth projections, and detailed studies of consumer preferences and behaviours.

Regulatory Compliance Data: Detailed records of each company's compliance history, including any breaches, penalties, or commendations, offering a comprehensive view of their legal and ethical standing.

Exploring Data

Financial Data Analysis: Conduct trend analyses, compare financial performances across different sectors, and examine the impact of economic policies on company finances.

Market Trend Analysis: Utilize advanced analytics to identify emerging market trends, assess sector vitality, and evaluate the impact of global economic shifts on local businesses.

Compliance Pattern Identification: Analyse compliance data to identify common legal challenges faced by businesses, understand the correlation between regulatory adherence and business success, and recognize industry-specific compliance trends.

Verifying Data Quality

Completeness: Rigorous assessment of data sets to ensure all critical information is included, especially in key areas such as revenue, market share, and legal compliance.

Consistency: Cross-reference data from multiple sources to ensure uniformity in how information is reported and recorded, paying special attention to standardizing financial figures and compliance terminology.

Timeliness: Use the latest data available to reflect the most current business and market scenarios.

Task 4. Planning your project

Hours spent on the project are spread evenly across both members for each task.

Data collection and cleaning (6 hours): Collect data from the chosen datafiles and clean the data to remove any missing values, outliers, and inconsistencies.

Data exploration (5 hours): Explore the data to identify patterns and relationships between different variables. For example, the relationship between a company's sector and its earnings or the relationship between a company's regulatory compliance and its market concentration.

Feature engineering (10 hours): Develop new features from the existing data, to improve the performance of the machine learning model. For example, develop new features by combining existing features or by extracting new features from the data.

Model development (30 hours): Develop a machine learning model to understand the factors influencing business success in Estonia and predict the likelihood of new businesses surviving or thriving based on current market conditions and historical data of business successes and failures.

Model evaluation (10 hours): Evaluate the performance of the machine learning model using appropriate metrics such as accuracy, precision, recall.

Documentation and Reporting (8 hours): Compile project documentation, create a summary report, and present findings.

Making a poster (8 hours): Making the poster with all the significant findings, providing a clear and overall understanding of the analysis and its implications.

Here are the methods and tools that we plan to use:

Python: We will use Python for data cleaning, data exploration, feature engineering, and model development.

Pandas: We will use Pandas for data manipulation and analysis.

Scikit-learn: We will use Scikit-learn for machine learning model development and evaluation.