

Financial Statement Analysis with Python and Income Forecasting Using Machine Learning

Abstract:

Financial statement analysis using the Python programming language enables a more comprehensive and detailed examination of a company's financial data. This approach includes balance sheet analysis, generation of profit and loss charts, calculation of net profit, and comparison of revenue with various expense categories. Additionally, by utilizing the company's annual income data, linear regression models are applied to predict revenue trends and analyze financial performance in the coming years. This integrated approach of data analysis and statistical forecasting provides a powerful tool for informed financial and strategic decision-making.

Author: Mohammad Tinati

Email: mohammad_tinati@yahoo.com

Introduction

In this section, I aim to theoretically discuss the role of artificial intelligence in financial analysis, and in the next section, I will focus on the details of the code written and the model used

Financial statement analysis is a fundamental activity in the process of financial management and decision-making within organizations. This analysis helps managers and decision-makers evaluate the company's past financial performance and make forecasts for the future. Given the complexity and rapid economic changes in today's business world, traditional methods of financial analysis may not fully address the precise and comprehensive needs of organizations. As a result, the use of modern information technology tools and advanced statistical models, particularly in financial data analysis, has become essential. In this context, Python programming language, as one of the most widely used tools for financial and economic data analysis, plays a key role due to its extensive capabilities in data processing and analysis, especially in handling large and complex datasets.

With powerful libraries like pandas, matplotlib, and seaborn, Python provides useful tools for financial data analysis, creating visual charts, and simulating financial trends. Not only is this language valuable for analyzing past data, but it can also be effectively used for forecasting future trends through the application of machine learning models. Forecasting company revenue in the coming years is one of the major challenges in financial management, and with the use of statistical modeling techniques such as linear regression, this task can be performed more accurately. These methods assist organizations in gaining a better understanding of their future financial condition and optimizing strategies based on this understanding.

In addition to revenue forecasting, another crucial aspect of financial statement analysis is identifying and evaluating costs and financial ratios, which directly impact the profitability and financial health of an organization. Analyzing financial ratios such as cost-to-income ratios, profit margins, and other key indicators provides a clear picture of a company's financial efficiency and performance,

which can guide managers in making strategic decisions. This research focuses on financial data analysis using Python tools and utilizes machine learning models to predict future income. It aims to provide a comprehensive framework for companies and financial analysts to improve decision-making processes and strategies in today's complex and ever-changing economic landscape.

Ultimately, this study seeks to leverage the capabilities of Python in data processing and advanced machine learning models for financial statement analysis and forecasting, offering a useful tool for financial analysts and decision-makers. This will help them predict future financial trends more accurately and take steps to optimize their financial performance.

Practical Section

In this practical section, I first focused on obtaining the data needed for financial analysis. The company can upload its historical financial data (including income and expenses) through an Excel file for further processing and result analysis. The data should consist of annual records, including income and expenses for each year, to allow for comprehensive financial analysis.

To facilitate the process, I implemented a Python script that allows the company to upload their financial data via an Excel file. The script processes this data and prepares it for further analysis, which includes generating relevant financial charts and performing statistical calculations.

The dataset used in this implementation consists of income and expense data for multiple years. The goal is to leverage this data to perform financial statement analysis, identify trends, and forecast future income based on historical information. This approach provides valuable insights for decision-making and future planning.

Next, I proceeded with the analysis by using Python libraries such as pandas for data manipulation, matplotlib for data visualization, and machine learning models, particularly linear regression, to predict the company's financial outcomes for future years. By examining the financial health indicators, we are able to uncover significant patterns and predict future performance, which is crucial for making informed business decisions.

After receiving the data, the program initially presents the income and expense results of previous years and future projections in the form of bar charts. These charts allow for a direct comparison between income and expenses for each year. Additionally, a separate chart displaying net income resulting from the company's gross profit is presented. This visualization enables the user to understand the profit, expenses, and net income for each year.

In the next section, I focused on the expense data, specifically working on improving the categorization of costs and handling outliers. After transferring the expense data to the code, outliers, which could arise due to human error or other discrepancies, are detected and displayed as a separate dataset for the user. This feature allows the analysis of the expenses, showing how much was spent on each category and identifying any significant deviations, such as a sudden increase in costs compared to previous years.

In the final section, my main focus was on predicting the company's future income using a linear regression model. After preprocessing and organizing the data to reduce errors and optimize the available data, I used the regression model to make predictions for the next year's income. Initially, I built the model manually without relying on libraries. Although this approach resulted in more precise predictions, it consumed a considerable amount of time when handling larger datasets. The model took a significantly long time to process. Moving forward, I plan to optimize the processing speed, as using available libraries can improve efficiency. However, I have observed that manually modeling the regression gives better results in some cases, though optimizing the time efficiency is an important consideration.

Linear Regression :

The diagram illustrates the linear regression equation $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$. Each term is labeled with an arrow pointing to it: Y_i is the Dependent Variable, β_0 is the Population Y intercept, β_1 is the Population Slope Coefficient, X_i is the Independent Variable, and ϵ_i is the Random Error term. A blue bracket under $\beta_0 + \beta_1 X_i$ is labeled 'Linear component', and another blue bracket under ϵ_i is labeled 'Random Error component'.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Limitations

Initial Limitation - Cost Segmentation:

Initially, we face a limitation in cost segmentation, which can be handled using the KNN model to achieve a more precise classification of each cost. By segmenting costs, we can gain a better understanding of the company's existing expenses, leading to more accurate final reports and analysis.

Limited Coverage of Financial Statements:

The second limitation is the lack of coverage for all financial statements, including wages, taxes, and value-added tax. To overcome this, each of these categories can be individually segmented. This process involves collecting all financial data of the company and analyzing it accordingly. Additionally, using a linear regression model is not recommended due to its higher probability of error. In my opinion, decision tree models or neural networks would be more suitable alternatives.

Access Restriction to Financial Software APIs:

The most significant limitation, in my opinion, is the lack of access and permission to use financial software APIs within the company. If we can gain access to a company's financial accounting software API and obtain permission to develop software, we can integrate all analytical charts directly into the software, eliminating speed limitations and enabling real-time access to financial data.

In this scenario, our model can be integrated into the software, allowing it to function as a combination of AI and financial software. Although it is possible to develop a separate and customized accounting software solution, this would require a considerable amount of time and a full development team.

Failure to consider external economic variables:

The current model does not take into account external economic factors that may impact the company's income and expenses. Factors such as inflation, exchange rates, or changes in government policies can significantly affect the company's financial outcomes, but these factors are not covered in the model.

Uncertainty of predictions:

Although the linear regression model provides predictions of income for future years, it assumes that future trends will be similar to historical data. However, financial predictions are inherently uncertain and can be influenced by unexpected conditions such as economic crises, changes in consumer behavior, or political instability. Therefore, the predictions of the model should be regarded as estimates rather than certainties for the future.

Failure to use more comprehensive financial data:

The analysis conducted focused solely on revenue and expenses, while other important financial indicators such as debt levels, assets, liabilities, and cash flows were not considered. A broader set of financial indicators could provide a more complete picture of the company's financial situation and increase the accuracy of predictions.

خلاصه مقاله

تحلیل صورت‌های مالی با استفاده از زبان برنامه‌نویسی پایتون، امکان بررسی دقیق‌تر و جامع‌تری از داده‌های مالی یک شرکت را فراهم می‌کند. این روش شامل تجزیه و تحلیل ترازنامه، ایجاد نمودارهای سود و زیان، محاسبه سود خالص و مقایسه درآمد با دسته‌بندی‌های مختلف هزینه است. همچنین، با استفاده از داده‌های سالانه درآمد شرکت، مدل‌های رگرسیون خطی برای پیش‌بینی روند درآمد و تحلیل عملکرد مالی در سال‌های آینده به کار گرفته می‌شوند. در این پژوهش تمرکز اصلی من برای کاهش خطای مدل رگرسیون و بهینه سازی داده ها میباشد. یکی از اساسی ترین مشکل های خطای مدل درست آنالیز و پردازش نکردن دیتا قبل از مدل سازی است که این امر باعث اشتباه در یادگیری ماشین میشود.

بخش نظری

تحلیل صورت‌های مالی، یکی از فعالیت‌های مهم در مدیریت مالی و تصمیم‌گیری سازمانی است که به مدیران کمک می‌کند تا عملکرد مالی گذشته شرکت را ارزیابی کرده و برای آینده پیش‌بینی کنند. با توجه به پیچیدگی و تغییرات سریع اقتصادی، روش‌های سنتی تحلیل مالی ممکن است پاسخگوی نیازهای دقیق و جامع سازمان‌ها نباشند. در این راستا، ابزارهای مدرن فناوری اطلاعات و مدل‌های آماری پیشرفته، به‌ویژه در تحلیل داده‌های مالی، ضروری شده‌اند.

پایتون به‌عنوان یکی از پرکاربردترین ابزارها در تجزیه و تحلیل داده‌های مالی و اقتصادی، امکانات گسترده‌ای برای پردازش و تحلیل داده‌ها ارائه می‌دهد. کتابخانه‌هایی مانند *pandas*، *matplotlib* و *seaborn* و *numpy* برای تحلیل داده‌ها، ایجاد نمودارهای بصری و شبیه‌سازی روندهای مالی بسیار مفید هستند. همچنین، از مدل‌های یادگیری ماشین مانند رگرسیون خطی برای پیش‌بینی درآمد شرکت‌ها در سال‌های آینده استفاده می‌شود.

اگرچه من در این کد از کتابخانه‌های دیگری هم استفاده نمودم اما در این بخش به مهمترین آنها اشاره کردم و کارایی هر یک با توجه به نیازهای ما متفاوت است. ابتدا سعی در ترکیب سازی کد بدون استفاده از کتابخانه نمودم این امر با زبان سی پلاس پلاس انجام دادم با توجه به افزایش

سرعت در دریافت تحلیل دیتا که بسیار برای من کارآمد بود ولی تصمیم براین گرفتم برای ذخیره زمان و برای رفع محدودیت های مدل از زبان پایتون که محبوب در این زمینه است استفاده بنماییم.

همچنین علاوه بر پیش بینی درآمد، یکی دیگر از جنبه های مهم تحلیل صورت های مالی، شناسایی و ارزیابی هزینه ها و نسبت های مالی است که بر سودآوری و سلامت مالی سازمان تأثیر می گذارند. این پژوهش بر روی تحلیل داده های مالی با ابزارهای پایتون و استفاده از مدل های یادگیری ماشین برای پیش بینی درآمد آینده متمرکز است. هدف این مطالعه ارائه یک چارچوب جامع برای بهبود فرایندهای تصمیم گیری مالی در محیط پیچیده و متغیر اقتصادی امروز است.

بخش عملی

در این بخش، ابتدا داده های مالی مورد نیاز برای تحلیل جمع آوری شده اند. شرکت ها می توانند داده های تاریخی مالی خود را (شامل درآمد و هزینه ها) از طریق یک فایل اکسل بارگذاری کنند. این داده ها باید شامل سوابق سالانه باشند تا تحلیل مالی دقیق تری انجام شود. یک اسکریپت پایتون برای پردازش و تجزیه و تحلیل این داده ها نوشته شده است که شامل موارد زیر است:

- رسم نمودارهای مالی برای نمایش روند درآمد و هزینه ها
 - استفاده از مدل های یادگیری ماشین (رگرسیون خطی) برای پیش بینی درآمد آینده
 - شناسایی ناهنجاری ها و هزینه های غیرعادی برای بررسی دقیق تر
- در نهایت، مدل رگرسیون خطی برای پیش بینی درآمد سال های آینده پیاده سازی شده است. این مدل ابتدا به صورت دستی بدون استفاده از کتابخانه های آماده اجرا شد که دقت بالاتری داشت، اما برای پردازش داده های بزرگ، زمان زیادی مصرف می کرد. در آینده، بهینه سازی این مدل برای افزایش سرعت پردازش مدنظر دارم.

محدودیت‌ها

1. محدودیت در دسته‌بندی هزینه‌ها : برای بهبود تحلیل هزینه‌ها، استفاده از مدل *KNN* پیشنهاد شده است. این روش به طبقه‌بندی بهتر هزینه‌ها و ارائه گزارش‌های دقیق‌تر کمک می‌کند.
2. عدم پوشش کامل صورت‌های مالی : برخی داده‌های مالی مانند حقوق، مالیات و ارزش افزوده به‌طور کامل در تحلیل پوشش داده نمی‌شوند. استفاده از مدل‌های پیشرفته مانند درخت تصمیم‌گیری یا شبکه‌های عصبی می‌تواند دقت پیش‌بینی‌ها را افزایش دهد.
3. عدم دسترسی به *API* نرم‌افزارهای مالی : یکی از مهم‌ترین محدودیت‌ها، عدم اجازه دسترسی به *API* نرم‌افزارهای حسابداری شرکت‌ها است. در صورت امکان دسترسی، می‌توان تمامی تحلیل‌های مالی را مستقیماً در نرم‌افزار حسابداری اجرا کرد و محدودیت‌های سرعت پردازش را برطرف نمود.
4. عدم در نظر گرفتن متغیرهای اقتصادی خارجی : مدل فعلی به عوامل اقتصادی خارجی که می‌تواند بر درآمد و هزینه‌های شرکت تأثیر بگذارد، توجه نمی‌کند. عواملی مانند تورم، نرخ ارز یا تغییرات در سیاست‌های دولتی می‌توانند نتایج مالی شرکت را به طور قابل توجهی تحت تأثیر قرار دهند، اما این موارد در مدل پوشش داده نشده‌اند.
5. پیش‌بینی‌های غیر قطعی : اگرچه مدل رگرسیون خطی پیش‌بینی‌هایی از درآمد در سال‌های آینده ارائه می‌دهد، اما این پیش‌بینی‌ها فرض می‌کنند که روندهای آینده مشابه با داده‌های تاریخی خواهند بود. با این حال، پیش‌بینی‌های مالی به طور ذاتی غیرقطعی هستند و می‌توانند تحت تأثیر شرایط غیرمنتظره‌ای مانند بحران‌های اقتصادی، تغییرات در رفتار مصرف‌کنندگان یا ناپایداری‌های سیاسی قرار بگیرند. بنابراین، پیش‌بینی‌های مدل باید به‌عنوان تخمین‌هایی در نظر گرفته شوند و نه قطعیت‌هایی برای آینده.
6. عدم استفاده از داده‌های مالی جامع‌تر : تحلیل انجام‌شده تنها بر روی درآمد و هزینه‌ها متمرکز بود، در حالی که دیگر شاخص‌های مالی مهم مانند سطح بدهی‌ها، دارایی‌ها،

تعهدات و جریان‌های نقدی در نظر گرفته نشدند. مجموعه وسیع‌تری از شاخص‌های مالی می‌تواند تصویر کامل‌تری از وضعیت مالی شرکت ارائه دهد و دقت پیش‌بینی‌ها را افزایش دهد.

عموماً هدف کلی من از این پروژه، ادغام هوش مصنوعی با حسابداری است، رویکردی که در آینده‌ای نه‌چندان دور، تحولی در حوزه مالی و اقتصادی ایجاد خواهد کرد. تحلیل صورت‌های مالی تنها محدود به پیش‌بینی آینده نیست، بلکه می‌تواند به‌عنوان یک ابزار هوشمند و تصمیم‌ساز، برای هر بخش از سیستم مالی پیشنهادات بهینه، استراتژی‌های کارآمد و تحلیل‌های عمیق ارائه دهد.

در مرحله نخست این پروژه، با استفاده از زبان برنامه‌نویسی پایتون، بخشی از مدیریت داده‌های مالی را طراحی و پیاده‌سازی نمودم تا سنگ‌بنای یک سیستم هوشمند تحلیلگر باشد.

با استفاده از شبکه‌های عصبی، می‌توان به پیشرفت‌های قابل توجهی در تحلیل داده‌های مالی دست یافت.

شبکه‌های عصبی به دلیل قدرت بالای خود در شبیه‌سازی مغز انسان، قادرند الگوهای پیچیده و غیرخطی را شبیه‌سازی کرده و پیش‌بینی‌هایی دقیق و قابل اطمینان از داده‌های مالی انجام دهند.