DATA SCRAPING AND PREDICTIVE ANALYSIS IN BUSINESS INTELLIGENCE

[1] Ms. Dhanya Anto, [2]Clinton,

[1] Guest Lecturer Computer Science, [2]Student
[1] Computer Science,
[1]Prajyoti Niketan College, Pudukad, Thrissur, Kerala-India

Abstract: Data is one of the most powerful and influential forces of our day. Data is important in practically every industry, from journalism to business to medical to politics. Predictive analysis has evolved from what was originally known as business intelligence. While business intelligence and predictive analytics both use similar concepts, business intelligence focuses on the past, and predictive analysis, as the name implies, focuses on the future. To make the marketing process more efficient, web scraping and predictive analytics can be combined. This can be accomplished through the use of a variety of techniques, such as business intelligence. Primary goal will be to make money for any business.

Index Terms - **Data scraping, predictive analysis, business intelligence**

I. INTRODUCTION

The application of various components of data science to create predictions about future occurrences or actions is known as predictive analytics. Organizations are using predictive analytics to help solve difficult problems and find new opportunities. Common uses are detecting fraud, reducing risks, healthcare and direct marketing etc. Historical and transactional data patterns can be used to identify future risks and opportunities. Predictive analysis has applications in a variety of sectors, including business. It is used to investigate and comprehend customer behavior, products, and a variety of other factors in order to determine risks and possibilities. Businesses must use predictive analytics to stay competitive. In other words, predictive analytics is a technique for analyzing data using business intelligence. If we want to spy on any current market firms, then the best option is to use a service of web scraping. Any firm, of course, evaluates its own services or products and keeps an eye on market trends. Understanding what your competitors are doing, on the other hand, is critical to staying in the game. As a result, web scraping for business intelligence plays a critical function in today's industry. Even minor details about your competition can help you outsmart them in the marketplace. Even minor details like staffing can reveal information about future business operations or competitor locations.

Data like price of items, price of stocks, different reports, market pricing and details of products, can be gathered with help of web scraping. The analysis discovering difficulties, resolving the availability of appropriate data, choosing which method can assist in finding a solution to an interesting topic, and communicating the result are all part of the process. The data must be separated into multiple processes for analysis, such as starting with its specification, assembling, organizing, cleaning, re-analyzing, using models and algorithms, and finally arriving at a final result. Business intelligence can also be utilized to assist in the making of business decisions. Data for business intelligence has been obtained manually over the years. The internet has made it feasible to access a large amount of data for the purposes of business

II. ABOUT WEB SCRAPING

The more data flowing around your business organization than ever, it's crucial to not only work to analyze data but also find tools that improve the efficiency and accuracy of your data so you can make decisions that don't affect your business. There are different types of data analytics tools but they all need a reliable source of quality data to function very efficiently. That means we need to find a solution for data collection that can help you get required data in huge volumes and in real-time. That easiest solution is by using a web scraping API. Web Scraping is an important technique used for extracting unstructured data from the websites and transforming that into structured data. Web scraping is like another form of data mining. The basic and important aim of the web scraping process is to mine information from different and unstructured websites and transform it into a comprehensible structure like spreadsheets, database or a comma-separated values (CSV) file. Web scraping is an important tool for gathering data for business intelligence. This is due to the fact that standard online scraping entails data collection, selection, and even pre-processing.

Business intelligence becomes a reality and a dynamic process thanks to web scraping. This is because the required business intelligence data may be retrieved via the internet via the web scraping procedure. Web scraping assists the predictive analysis process by providing historical data that can be evaluated and predictions of consumer behaviours, such as customers who are likely to buy, renew, or buy similar products. For any corporate marketing effort, predictive analysis and site scraping are critical. Because marketing is an investment, it is vital for firms to use web scraping to obtain the necessary data for making business decisions. Predictive analysis narrows your target market and allows you to target individual clients with your marketing. This allows the market teams to create a variety of adverts based on the data. Because web scraping is a vital aspect of predictive analysis, it's crucial for a business to invest in the process. Companies must engage clients who are more inclined to respond positively. Marketing tactics will only be effective if a company can target goods and services that clients require at the right moment. A web scraper is an API that allows you to extract data from a website. Predictive analytics can help you save money by lowering the amount of money you spend to make a transaction.

III. WHAT IS PREDICTIVE ANALYSIS?

Predictive analysis is the practice of analyzing existing data to find patterns and predict future outcomes or trends. Predictive analysis cannot predict the future with certainty, but it can foresee the odds. Predictive analysis has applications in a variety of sectors, including business. Predictive analysis is used to investigate and comprehend customer behavior, products, and a variety of other factors in order to determine risks and possibilities. However, as is clear, it is a type of analysis that is carried out using a large amount of existing data. Businesses must use predictive analytics to stay competitive. Predictive analytics is a way or a technique for examining quality data using business intelligence. This is due to the fact that it is employed in forecasting and modelling. It is a pattern-prediction approach with numerous applications in the credit, medical, and insurance industries. Credit evaluation is the most typical application of web scraping and predictive analytics integration.

The utilization of historical data to forecast a company's and market's future is a crucial aspect of any organization. Business intelligence is critical in assisting marketing teams in anticipating and preparing for client needs rather than reacting to them. Data based on demographics that may have been neglected in the past can be presented through web scraping. Any combination of client demographics can be valuable in determining which marketing platform to employ, as well as which marketing strategy to use and when. Managers can use a combination of site scraping and predictive analytics to increase sales while spending less money. Any business's goal is to reduce losses and also increase earnings. As a result, web scraping and predictive analysis are critical for businesses, whether they are online or offline.

IV. WEB SCRAPING TECHNIQUES AND METHODS:

Steps involved in web scraping:

Find the URL of the webpage that you want to scrape
Select the particular elements by inspecting
Write the code to get the content of the selected elements
Store the data in the required format.

V. POPULAR LIBRARIES OR TOOLS USED FOR WEB SCRAPING:

Selenium – a framework for testing web applications
BeautifulSoup – Python library for getting data out of HTML, XML, and other markup languages
Pandas – Python library for data manipulation and analysis

We are going to use the Python module requests and BeautifulSoup. The requests will allow you to send the HTTP requests to get the HTML files. BeautifulSoup will be used to parse the HTML files. It is one of the most used libraries for web scraping. It is very simple to use and has so many features that help gather websites data efficiently.

Prerequisites

A. python 2.7
B. requests
C. beautifulsoup4
D. pandas

These are some other best scraping tools available today:

E. ParseHub
F. Scrapy
G. OctoParse
H. Scraper API
I. Mozenda
J. Webhose.io
K. Content Grabber
L. Common Crawl

VI. APPLICATIONS OF WEB SCRAPING IN BUSINESS:

Web Scraping helps you to gather data for testing or training. When the data is not available we can use web scraping to collect data from various websites.

Some of the most common uses of web scraping:

1. Scrap product details from ecommerce websites to show on own websites, to provide price comparisons. Companies that provide products or services must have detailed information on competitors' products and services, which arrive on the market on a daily basis. To maintain a constant eye on this data, web scraping tools might be used.

2. Real estate- Web scraping software can be used to collect property details from real estate websites. Web scraping can be used to scrape agent and owner contact information in addition to property details.

3. Marketing Lead Generation-
To create leads for marketing, web scraping tools can be used. Scraping data from relevant websites can be used to create email and phone lists for cold outreach. Business contact information, such as phone numbers and email addresses, can be scraped from yellow pages websites or Google Maps business listings.

4. Extract profile data from social networks.

5. Scrap hospital/clinic websites to build a catalog of health physicians including their contact details.

6. Scrap reviews of products, places and hotels or restaurants etc.

7. E-commerce-   Web scraping is a technique for extracting product data from e-commerce websites such as Amazon, eBay, and Google Shopping on a regular basis. A web scraping software can easily extract product details such as pricing, description, photos, reviews, and ratings.

VII. TOOLS FOR PREDICTIVE ANALYTICS

Predictive analytics software provides customers with detailed, real-time insights on an almost infinite number of business activities. Predictions based on an analysis of data collected over a period of time can be used to find out many types of behavior and patterns.

a. Tools for descriptive data analytics:

One of the most fundamental types of analysis is descriptive analysis. These analytics depict historical occurrences, no matter how recent or distant they were, after the raw data has been processed. This is beneficial because firms can use this information to learn from past actions and make better ones in the future. This is usually the starting point for most quantitative and visual data analysis. Data analysis on a company's customer base, production information, financial records, and sales are all examples of descriptive analytics.

b. Predictive data analytics tools

Predictive analysis accomplishes precisely what its name implies: it forecasts future events. While the future can never be completely predicted, you can predict the likelihood of a variety of distinct scenarios. This enables businesses to consider risk when making decisions, effectively putting these probabilities on their odds. Forecasting supply trends to make decisions on what kind of supplies to utilize and where they're acquired from, or predicting your sales revenue in the next few years, are examples of predictive analysis.

c. Diagnostic analytics tools

The most abstract phase of analysis is sometimes referred to as diagnostic analysis. It explains why some occurrences occur, or why specific regions of manufacturing were in such short supply or great demand. Recognizing correlations between occurrences and determining whether they have a positive, negative, or zero correlation is also part of diagnostic analysis. When both variables rise at the same time, there is a positive association. When one variable rises while the other falls, this is known as a negative correlation. The term "zero correlation" refers to the fact that neither the variables are related to one another, nor are they impacted by one another.

d. Tools for prescriptive data analysis

Prescriptive analysis is becoming increasingly common in businesses of all sizes. Prescriptive analysis entails combining all of the aforementioned data analysis methodologies and providing various possibilities for a specific business decision. It is possible to anticipate the implications of these actions and prospective outcomes, as well as to explain why these results will occur.

It will accurately forecast occurrences, explain why they might occur, and offer options for making business decisions based on these events.

Many predictive analytics users rely on tools created by third-party developers. Almost all of the solutions are specifically built to fit the demands of some businesses and departments. The following are some of the most well-known predictive analytics software and service providers:

I. Acxiom
II. IBM
III. Information Builders
IV. Microsoft
V. SAP
VI. SAS Institute
VII. Tableau Software
VIII. Teradata
IX. TIBCO Software

VIII. MODELS OF PREDICTIVE ANALYTICS

Predictive analytics is built on models, which allow users to turn historical and present data to find insights, resulting in beneficial long-term outcomes. Predictive models come in a variety of shapes and sizes.

1. <u>Customer Lifetime Value Model</u>: Finds out the customers who are most likely to invest more in products and services.
2. <u>Customer Segmentation Model</u>: Group customers based on similar characteristics and purchasing behaviors
3. <u>Predictive Maintenance Model</u>: Forecast the chances of essential equipment breaking down.
4. <u>Quality Assurance Model</u>: Find and prevent defects and extra costs when providing services to customers

IX. PREDICTIVE MODELING TECHNIQUES

Model users have access to an almost endless range of predictive modelling techniques. Many methods are unique to specific products and services, but a core of generic techniques, such as decision trees, regression and also neural network which are now widely supported across a wide range of predictive analytics platforms.

Decision trees- one of the most popular techniques, rely on a schematic, tree-shaped diagram that is used to determine a series of action or to show a probability statistical. The branching method can also show every possible outcome of a particular decision and how one choice may lead to the next. The branching approach can also indicate all of the various outcomes of a given decision, as well as how one choice may lead to the next.

Banking, investing, and other finance-related models frequently employ **regression techniques.** Users can use regression to forecast asset values and understand correlations between variables like commodities and stock prices.

Neural networks are at the cutting edge of predictive analytics techniques, as they are algorithms that imitate the way the human mind works to uncover underlying links within a data set.

X. PREDICTIVE ANALYTICS ALGORITHMS

Predictive analytics adopters can quickly access a variety of statistical, data-mining, and machine-learning methods for use in predictive analysis models. Algorithms are typically created to tackle a specific business problem or set of problems, improve upon an existing algorithm, or provide a new capacity.

Customer segmentation, community recognition, and other social-related tasks, for example, are well-suited to clustering algorithms. Classification algorithms are commonly used to boost customer retention or construct a recommendation system. When creating a credit rating system or predicting the result of several time-driven events, a regression algorithm is usually chosen.

XI. SOME APPLICATIONS OF PREDICTIVE ANALYSIS:

Price comparisons with competitors

Price is extremely important in today's e-commerce world. Keeping an eye on competitors' pricing will be helpful.
Trying to keep track of prices by hand, on the other hand, is not a practical choice. Furthermore, rates fluctuate from time to time. As a result, manually keeping track of prices becomes nearly impossible.
This is when web scraping is important. It automates the process of obtaining competitor prices and keeps you informed about new pricing   techniques used by your competitors.

XII. CONSUMER SENTIMENT ANALYSIS

Consumer attitudes must be monitored and analyzed. It's done by looking at customer comments and evaluations of various businesses. However, manually collecting all of the reviews from various sources is not practicable. As a result, web scraping is used to make it incredibly simple. Web scraping allows you to compile all of the reviews into a spreadsheet and compare them based on keywords.

XIII. REAL-TIME ANALYTICS

Real-time analytics essentially means analyzing data as soon as it becomes available. It differs from batch-style analytics in that it can take hours or even days to process data and generate insights. In comparison, real-time analytics can provide information immediately. Credit scoring is done using real-time analytics by financial organizations to determine whether to extend or terminate credit.
Customer relationship management (CRM) is an excellent illustration of how real-time analytics can be used to improve customer happiness and company outcomes.

XIV. PROCESS OF INTEGRATING SCRAPED DATA INTO PREDICTIVE ANALYTICS TOOLS

API is a visual interface that allows software pieces to connect. Web scraping allows you to collect data from any website online using bots known as "crawlers" or "spiders". These bots will analyze the whole website source code and get the data based on predetermined parameters. Web scraping software allows you to capture as much data as you want.

Using the web scraping API, you can link between your scratch software and your data analysis software. This allows you to send data from a web scraper directly to your analytics software. The Internet scraping API captures data from source, scraping software, and sends it to your data analysis tools directly. This eliminates the difficulty of collecting data, compiling it, formatting it, and finally installing it in your data analysis software.

Web scraping data can be used to extract competent product prices and inventories and compare them with your own, allowing you to do descriptive analysis and better understand how your company works. You can get real-time price updates using the scraping API. The API sends a scratch request whenever the price changes and provides updated data to your data analysis tools. In the event that your company makes a marketing attempt to attract a new targeted customer group that a competitor already has, a mock API may be used to conduct forensic research. You can collect information from your target customers based on their social media profiles using the scraping API. By releasing these profiles, you can identify trends and do forecasting research to better understand and build a marketing campaign for a new group of consumers.

Scattering reviews of your business is one way to apply scratching to your diagnostic analysis to see why people are talking about your company and why. This may be due to Scraping Robot's scraping API. Gather information about your company's aggregated estimates, reviews, occupational health balance, pay, security, management, culture, and written reviews. All you have to do is fill in the name of your organization, and details will be collected and sorted by you! After you have collected this information, you can begin your diagnostic research to find out what your company is doing right and wrong, and make changes to help it grow and prosper.

By recognizing that data obtained from all other mentioned types of analysis are used in the mean analysis, the opportunity for clarity comes by combining all previously extracted data. The scraping API allows you to compile your data collection and create data feeds that are not directly fed into your data analysis tools. Setting up a scraping panel from scraping software to API to data analysis tools allows you to gather all your data in one place instead of waiting for your company's various departments to deliver data from multiple sources.

XV. WEB SCRAPING IS ALSO USED IN OTHER INDUSTRIES.

News and Reputation Monitoring News

Scraping is quite beneficial for keeping track of information about individuals, products, and companies. Web scraping is an important part of extraction of data in the form of news from a variety of sources. After that, the data can be analyzed to get the necessary outcomes. As a result, it is also possible to maintain track of a company's brand and reputation.

Academic

Academic field depends a whole lot on data. Web scraping has now made it so easier for them to extract their needful data.

Data Journalism

The reason why data matters to them a lot is because data provides credibility to the arguments and claims made in the stories. Web scraping comes in here because it makes the data available and enables the journalist to create the impact through the creative use of the data.

Non-Profit

They need data in order to define their mission and further their work. Web scraping tools make it easy to extract data they need to reach their goals.

Employment

The Job Board uses crawlers to crawl different websites and scrape the information regarding new job postings. Job Boards collects information regarding job postings, job descriptions, and employee profiles. This enables them to provide information like job postings and connect job seekers with employers. Web scraping makes it all possible.

XVI. FUTURE AND SCOPES

Business intelligence can be thought of as a decision support system in which data is gathered for predictive analysis. It can also be utilized to assist in the making of business decisions. Data for business intelligence has been obtained manually over the years. The internet has made it easy to access a huge amount of data for the purposes of business. Businesses must use predictive analytics to stay competitive. Predictive analytics is a method for examining data with the help of business intelligence.

The Internet of Things (IoT) will grow enormously. Analytics tools are dealing with the massive amounts of structured and unstructured data generated by IoT. These are used for predictive analysis, usually used to examine the historical performance of the organization. Business analytics looks at reports generated by business intelligence (BI) to an algorithm that is able to make choices for us. Predictive and prescriptive analytical approaches will make many companies rethink traditional models favoring old technologies that increase sustainability and chances of success.

XVII. CONCLUSION

As the Internet has grown enormously and businesses have become increasingly reliant on data, having access to the most up-to-date information on any particular topic has become essential.

Data has become the core of all decision-making processes, whether in a business or a non-profit. As a result, web scraping is now used in practically every significant endeavor in the modern world.

It's also becoming clear that individuals that use a web scraping tool in a new and unique way can get a leg up on the competition and get a competitive advantage. Business intelligence and predictive analytics are fundamentally different in that business intelligence tries to answer questions like "what is happening now," whereas predictive analytics takes a more practical approach to assessing data. Raw data is converted into information for direct use by humans in the business intelligence process. In the case of predictive analytics, the raw data is transformed into structured data for the organization's productive usage and future projections.

Decision - Business intelligence aids people in making decisions based on information, whereas predictive analytics aids firms in making fact-based judgments.

XVIII. REFERENCE:

1 https://www.loginworks.com/blogs/predictive-analytics-web-scraping/?amp

2. "Predictive Analytics in Data Science for Business Intelligence Solutions"-(Parth Wazurkar , Robin Singh Bhadoria ,Dhananjai Bajpai)

3. https://www.octoparse.com/blog/using-web-scraping-to-improve-business-analytics-and-intelligence#

4. https://oxylabs.io/blog/data-mining-vs-machine-learning

5. https://www.cio.com/article/3273114/what-is-predictive-analytics-transforming-data-into-future-insights.html

6. https://inpressco.com/wp-content/uploads/2015/03/Paper49856-860.pdf

7. "Web scraping with Python: Collecting more data from the modern web" -Ryan Mitchell

8. https://www.cio.com/article/2439504/business-intelligence-definition-and-solutions.html

9. https://olap.com/business-intelligence-versus-predictive-analytics/

10. https://www.logianalytics.com/predictive-analytics/what-is-predictive-analytics/

11. https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/

12. https://www.cio.com/article/3273114/what-is-predictive-analytics-transforming-data-into-future-insights.html

13. https://www.upgrad.com/blog/scope-of-business-analytics-in-india/

14. https://www.webharvy.com/articles/web-scraper-use-cases.html

15 https://scrapingrobot.com/blog/data-analysis-tools/

16. https://www.promptcloud.com/blog/data-mining-and-predictive-analysis/