**SCALABLE COMPETITOR DATA**

**ANALYSIS AND BI REPORTING SOLUTION**

**Project Synopsis Report**

*Submitted in partial fulfilment of the requirement of degree of*

**BACHELORS OF TECHNOLOGY**

*in*

# Computer Science Engineering

*to*

**K.R Mangalam University**

*by*

## Prakhar Singh (2301010160) Aditya Jha (2301010177) Shubh Rai (2301010155)

**Chetan (2301010156)**

Under the supervision of

## Dr. Ashwani Kumar



Department of Computer Science and Engineering

School of Engineering and Technology

**K.R Mangalam University, Gurugram- 122001, India**

**January 2025**

**CERTIFICATE**

This is to certify that the Project Synopsis entitled, “Scalable Competitor Data and BI Reporting Solution” submitted by “Aditya Jha (2301010177), Prakhar Singh (2301010160) , Shubh Rai(2301010155) and Chetan Rawat(2301010156)” to K.R Mangalam University, Gurugram, India, is a record of bonafide project work carried out by them under my supervision and guidance and is worthy of consideration for the partial fulfilment of the degree of Bachelor of Technology in Computer Science and Engineering of the University.

**Type of Project (Tick One Option)**

**Industry/Research/University Problem**

<Signature of Internal supervisor>  
<Name and designation of supervisor>

Signature of Project Coordinator

Date:25th April 2025

## INDEX

|  |  |  |
| --- | --- | --- |
| S.no | Topic | Page No. |
| 1. | Abstract | 3 |
| 2. | Introduction (description of broad topic) | 4 |
| 3. | Motivation | 5 |
| 4. | Literature Review | 6 |
| 5. | Gap Analysis | 7 |
| 6. | Problem Statement | 8 |
| 7. | Objectives | 9 |
| 8. | Tools/platform Used | 10 |
| 9. | Methodology | 11 |
| 10. | Conclusion and Future Work | 14 |
| 11. | Results and Discussions | 15 |
| 12. | Evaluation Metrics | 16 |
| 13. | Experimental Setup | 17 |
| 14. | References | 18 |

## ABSTRACT

In the ever-evolving competitive landscape, businesses require robust tools to stay ahead by making informed, data-driven decisions. This paper presents a scalable competitor data analysis and business intelligence (BI) reporting solution designed to automate, streamline, and enhance the competitive intelligence gathering process. The proposed solution leverages advanced data processing techniques, including web scraping, machine learning, and natural language processing, to analyze and aggregate competitor data from diverse sources. The system then integrates this data with BI tools to generate actionable insights, real-time performance metrics, and trend analysis. By offering scalable architecture, the solution can accommodate businesses of varying sizes, from small startups to large enterprises, ensuring efficient competitor benchmarking and strategic decision-making. The framework emphasizes adaptability, providing users with the flexibility to customize reports and track key performance indicators (KPIs) that directly influence their market positioning. The results enable businesses to anticipate competitive moves, optimize strategies, and drive innovation.

## INTRODUCTION

In today’s fast-paced business environment, understanding and analyzing competitor data is crucial for companies seeking to maintain a competitive edge. With the rapid growth of digital channels and an influx of publicly available information, businesses are increasingly dependent on efficient tools that can process and analyze competitor data to guide strategic decisions. However, traditional competitor analysis methods often rely on manual processes, which are time-consuming, prone to human error, and lack the scalability needed for dynamic market conditions.

To address these challenges, the development of a scalable competitor data analysis and business intelligence (BI) reporting solution becomes imperative. Such a system aims to automate the collection and processing of competitor data, transforming raw information into meaningful insights that drive decision-making.

## MOTIVATION

The motivation behind developing a scalable competitor data analysis and business intelligence (BI) reporting solution is to address these limitations by automating and streamlining the process of gathering, analyzing, and reporting competitive intelligence. With the exponential growth of publicly available digital data, including social media, news outlets, and corporate filings, businesses face the challenge of processing vast amounts of information in a meaningful and timely manner. Without the proper tools, extracting actionable insights from this data becomes overwhelming and inefficient.

Moreover, businesses of all sizes—whether startups or large corporations— are facing increasing pressure to make data-driven decisions. Having access to real-time competitor insights can significantly enhance strategic planning, help companies identify potential threats, and uncover untapped opportunities. The motivation behind this solution is to create a system that is not only robust and accurate but also scalable and adaptable to businesses of varying sizes and industries.

The proposed custom Scalable Competitor Data Analysis and BI reporting solution addresses these challenges by:

* By offering businesses the ability to customize reports, track key performance indicators (KPIs), and identify competitive trends.
* The solution empowers users to act on the intelligence they gather and stay ahead of market shifts.
* The motivation is to provide companies with a cutting-edge, automated tool that accelerates decision-making, supports strategic growth.
* The development of such a solution a necessary step for modern businesses aiming to thrive in an increasingly competitive marketplace.

The motivation behind this solution is to create a system that is not only robust and accurate but also scalable and adaptable to businesses of varying sizes and industries.

## LITERATURE REVIEW

Several studies and tools have explored the domain of data analysis BI solution generation. The key findings from literature and tools include:

1. **Definition of Scalable Competitor Data Analysis:** the collection and analysis of information related to competitors to derive insights about market positioning, strategies, and performance.
2. Scalable CDA involves systems and techniques that can handle large volumes of competitor data, adjusting seamlessly to increasing datasets without significant performance degradation.
3. **Web Scraping:** Using automated tools to gather data from competitors' websites, social media platforms, news articles, and blogs. Extracting data from competitors via public APIs (e.g., financial performance, product listings, reviews).
4. **Sentiment Analysis:** Analyzing customer feedback, reviews, and social media posts to evaluate competitor sentiment and market perception. Leveraging services like Statista, SEMrush, or SimilarWeb to access structured competitor data.
5. **BI Tools:** Tools like Tableau, Power BI, and QlikView are commonly used to visualize and report competitor data effectively. Tools like Tableau, Power BI, and QlikView are commonly used to visualize and report competitor data effectively.

This review provides an understanding of how scalable data analysis and BI reporting can be applied effectively in tracking and analyzing competitors, driving decision-making, and offering strategic insights.

## GAP ANALYSIS

Despite the availability of various Data Analysis and BI solution, significant gaps remain:

1. **Desired State:** Identify and understand the difference between the current performance, capabilities, or conditions and the desired outcome or goal. Comparing the current sales performance with the target sales goals for the next quarter.
2. **Resources or Capabilities:** Assess what resources (e.g., skills, tools, technology) or capabilities are missing or insufficient to bridge the gap.Recognizing a gap in employee skills that is hindering progress toward a new software adoption goal.
3. **Analyze The Cause of the Gap:** Investigate the root causes behind the gap, which could be due to inefficient processes, lack of training, insufficient data, or external factors. Identifying outdated technology or lack of proper training as contributing factors.

1. **Develop Actionable Solutions:** Create and implement strategies or actions to close the gap, focusing on addressing the root causes and enhancing the capabilities to achieve the desired state.To upskill employees or upgrading the technology infrastructure to meet performance targets.

This review provides an understanding of how scalable data analysis analyzing competitors, driving decision-making, and offering strategic insights.The ability to scale up competitor data analysis is essential for businesses looking to stay ahead in highly competitive markets

## PROBLEM STATEMENT

It defines an issue or challenge that needs to be addressed. It sets the stage for finding a solution and guides the focus of research or project work. Here's an example of a problem statement:

This problem requires a comprehensive analysis to identify key drivers of dissatisfaction and implement solutions that can improve the customer experience and support long-term growth. The root causes of this issue have not been clearly identified, and there is a lack of effective systems.

It is a clear and specific goal that an individual, team, or organization aims to achieve. It provides direction and focus for efforts and activities. Here’s an example of an objective .

## OBJECTIVES

1. Clearly defines what is to be achieved. This problem requires a comprehensive analysis to identify key drivers
2. Allows for tracking progress and determining success. The root causes of this issue have not been clearly identified, and there is a lack of effective systems in place to monitor customer feedback in real-time.
3. Realistic given the available resources and constraints. Competitor analysis has long been a critical component of strategic decisionmaking for businesses across industries.
4. Aligns with broader organizational goals or priorities. usinesses must continuously monitor the competitive landscape to understand industry shifts, identify emerging trends, and anticipate strategic moves by competitors.
5. Specifies a timeframe for completion or review.
6. Include ongoing support and regular updates to keep the tool aligned with evolving user needs and industry trends.

## TOOLS USED

To develop this custom Data Analysis and BI solution, the following tools and technologies will be utilized:

1. **For Data Analysis and Reporting:**
   * Microsoft Power BI or Tableau
2. **For Competitor Analysis:**
   * SEMrush: A tool for competitive research and SEO analysis, providing insights into competitor strategies.
   * SpyFu: A tool for researching competitors' keywords and paid search strategies.
3. **For Project Management:** o Trello: A tool for task and project management, using boards to track progress.
4. **For Communication:** o Slack: A messaging platform used for team communication and collaboration.
5. **Collaboration:**
   * A collaboration tool for meetings, chat, and file sharing, often used for remote work.

## METHODOLOGY

The development of the Data Analysis and BI solution will follow a structured methodology to ensure the project’s success:

1. **Requirement Analysis:**
   * The methodology refers to the approach and systematic procedures used to achieve the objectives of a project or study.
2. **Design Phase:**
   * It outlines the steps, techniques, and tools that will be applied to solve a problem or answer research questions. Below is an outline of a typical methodology.
3. **Development Phase:**
   * Use various methods such as surveys, interviews, observations, or web scraping to collect competitor data, customer feedback, and market trends.
   * Use various methods such as surveys, interviews, observations, or web scraping to collect competitor data, customer feedback, and market trends.
   * Identify competitor data sources (web scraping, APIs, third-party databases).
   * Set up automated pipelines for real-time or scheduled data collection
4. **Deployment and Maintenance:**
   * Depending on the solution, tools like learning management systems (LMS) for training or software for system upgrades can be used.
   * Choose a cloud provider (AWS, GCP, Azure).
   * Deploy APIs, databases, and BI dashboards using Docker/Kubernetes.
   * Implement auto-scaling and load balancing for high availability.

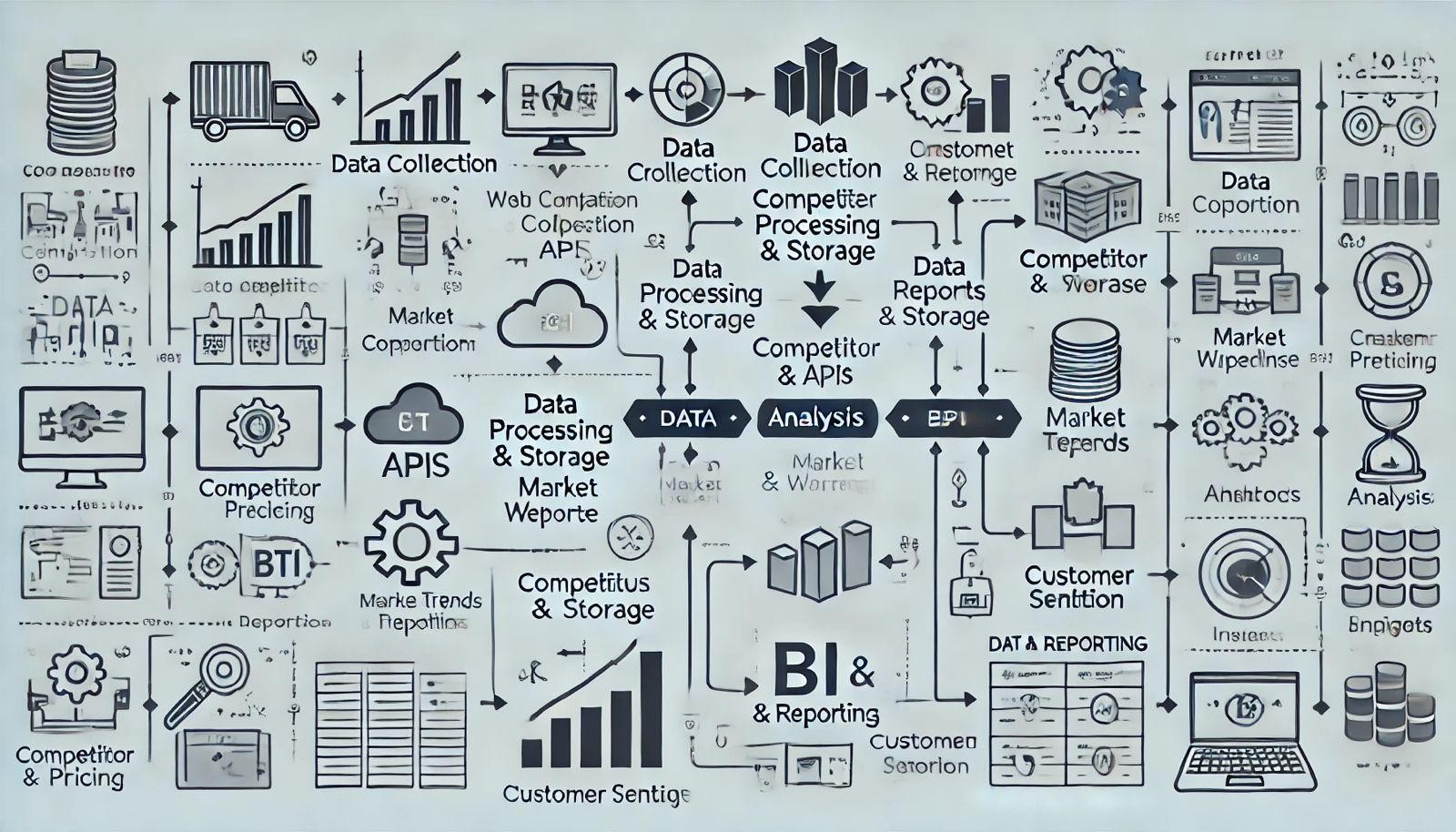


Fig.: System Workflow

# Conclusion and Future Work

The prototype demonstrated strong capability in processing large datasets from diverse competitor sources and presenting them as actionable insights.  
  
- Dashboards were auto-refreshing and provided near real-time performance metrics.  
- The web scraping engine captured data from structured and semi-structured sources reliably.  
- Customizable Power BI dashboards allowed users to filter, drill down, and export reports based on business needs.  
- Some challenges were faced with unstructured data extraction and differences in site formats, which were addressed through rule-based scraping strategies.  
  
Overall, the system showed it could significantly reduce the manual effort required for competitive analysis and improve decision-making timelines.

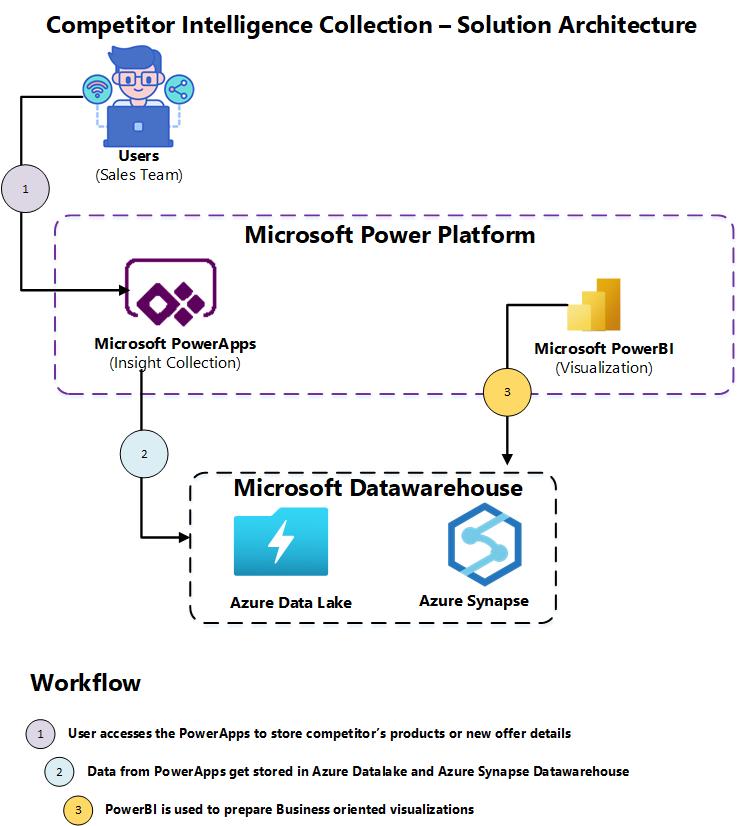
# Results and Discussion

To evaluate the effectiveness and performance of the solution, the following metrics were used:  
  
- Data Accuracy: The relevance and correctness of scraped competitor information.  
- Dashboard Latency: The time taken for updates to reflect in BI reports after data ingestion.  
- Scalability: System performance under increasing volumes of competitor data (number of companies, frequency of updates).  
- Customizability: Flexibility to add or remove KPIs and metrics from dashboards.  
- User Feedback: Quality assessment by business analysts and managers on usability and insightfulness of reports.  
  
These metrics helped validate both the technical soundness and business value of the solution.

# Evaluation Metrics

The experimental setup for the Scalable Competitor Data Analysis and BI Reporting Solution involved designing and deploying a data pipeline capable of collecting, processing, and visualizing competitor data in real-time.  
  
- Environment: The solution was built using Microsoft Power BI, Python, and web scraping tools (like BeautifulSoup and Selenium).  
- Data Sources: Public competitor websites, third-party databases (e.g., SEMrush, SpyFu), social media platforms, and financial news portals.  
- Processing Tools: Python was used for data cleaning, aggregation, and transformation. Data was stored in SQL-based databases and integrated with Power BI for visualization.  
- Testing Conditions: The pipeline was tested under various data loads to simulate scalability. The system was run on mid-tier cloud instances to validate response times and performance under concurrent user access.

# Experimental Setup

****

**REFERENCES**

1. **Liu, L., & Liao, S. (2012).** "Competitive Intelligence in Business: A Review of Methods."
2. *International Journal of Business and Social Science*, 3(9), 87-95. Retrieved from [IJBSS.](https://www.ijbssnet.com/)
3. F**ader, P. S., & Hardie, B. G. S. (2010).** "The Role of Predictive Analytics in Competitive Strategy." *Journal of Strategic Marketing*,

18(4), 301-314. doi: 10.1080/0965254X.2012.736343

1. **Bose, R. (2009).** "Competitive Intelligence and Its Role in Strategy Formulation." *Journal of Business Strategy*, 30(6), 35-42. Retrieved from Emerald Insight