A PROJECT REPORT ON

**SENTIMENT INTENSITY ANALYSIS OF GOOGLE REVIEWS ON EDUTECH COMPANIES IN KOCHI**

*Submitted By*

**ABEL JACOB VARGHESE**

**CHAPTER 1**

**INTRODUCTION**

In today’s digital-first world, student reviews and feedback have become a critical factor in shaping the reputation and effectiveness of educational institutions. The Edu Tech sector, especially in cities like Kochi, has seen rapid growth in recent years with numerous institutes offering short-term and long-term courses in IT and emerging technologies like Python, Data Science, Artificial Intelligence, and more.

This project, titled “Sentiment Intensity Analysis of Google Reviews on Edu Tech Companies in Kochi,” aims to derive actionable insights from publicly available reviews using data science techniques. By analysing both the ratings and the textual sentiments expressed in reviews, we can better understand student satisfaction, expectations, and areas for improvement for these institutions.

What makes this project unique is the introduction of a custom feature called “Stack of Learning”—a column that extracts technology-specific keywords (e.g., Python, Java, ML, etc.) from each review. This helps in understanding which courses are most talked about and how learners feel about them specifically.

The project involved several stages of execution, including:

* Data Collection from platforms like Google Reviews using automated web scraping.
* Data Cleaning and Preprocessing to handle noise, duplicates, and formatting inconsistencies.
* Sentiment Analysis, where we evaluated user sentiment both from star ratings and natural language content.
* Data Visualization, where insights were presented in a dynamic and interactive Power BI dashboard.
* Time Series and Stack-wise Analysis, allowing trends and performance of courses and institutes over time to be monitored.

Through this project, the goal is to showcase not only the technical application of NLP and data analysis but also how such methods can be used by Edu Tech companies, training centres, and even learners to make more informed decisions. The results can support institute-level strategic planning, marketing, and curriculum enhancement based on real, unfiltered student voice.

This project is part of my portfolio to demonstrate practical application of Python, Web Scraping, Sentiment Analysis, Power BI, and Data Storytelling, and contributes to my journey towards becoming a professional Data Scientist.

**CHAPTER 2**

**2.1 OBJECTIVES**

* To collect and clean review data from multiple online platforms.
* To identify frequently mentioned technologies and courses (Stack of Learning).
* To perform sentiment analysis based on both textual content and user ratings.
* To visualize trends, polarity, and review distribution through interactive dashboards.
* To derive insights for improving Edu Tech service quality and user satisfaction.

**2.2 TOOLS & TECHNOLOGIES USED**

* Python – for scripting and sentiment analysis
* Selenium – for web scraping with dynamic scrolling and multi-page support
* BeautifulSoup(bs4) – for parsing and extracting structured data from HTML
* Pandas – for data manipulation and cleaning
* NLTK / TextBlob / VaderSentiment – for sentiment polarity classification
* Power BI – for data visualization and interactive dashboards
* GitHub – for code versioning and project documentation

**CHAPTER 3**

**3.1 DATA COLLECTION**

Reviews were scraped using Selenium automation from Google Maps, simulating scrolling and dynamic content loading. Data was gathered from selected EdTech institutions in Kochi.

Collected Attributes:

* Review ID, Name, Text
* Star Rating
* Date of Review
* Pre-processed Review
* Detected Stack of Learning (keywords)

Each review was tagged using a dual-layer sentiment mapping approach:

* Star Rating (1–5 scale)
* Text Polarity via NLP (TextBlob)

This combination helped resolve inconsistencies like overly positive text with low ratings. A total of 1,000+ reviews were collected, covering mostly positive sentiment, with a fair spread of neutral and negative ones.

**3.2 DATA PREPROCESSING**

* Removed duplicates and filtered non-text entries.
* Lowercased and cleaned the review text.
* Removed emojis, special characters, and stop words.
* Applied polarity and subjectivity scoring using TextBlob.
* Tagged course-specific keywords using pattern matching.
* Created two new columns: Rating Sentiment and Text Sentiment.

**CHAPTER 4**

**SENTIMENT ANALYSIS**

Sentiment analysis in this project was enriched through a dual-layered approach—leveraging both star ratings and natural language polarity from review text. This ensures more accurate representation of user sentiment, especially when users leave contradictory reviews such as high ratings with negative remarks or vice versa.

We used polarity scoring (TextBlob) where text sentiment was classified as:

• Positive (polarity > 0)

• Neutral (polarity = 0)

• Negative (polarity < 0)

And for Rating Sentiment:

• Positive: 4–5 stars

• Neutral: 3 stars

• Negative: 1–2 stars

Combining both metrics helped in surfacing anomalies and offered a way to verify inconsistencies in user feedback. Furthermore, mismatches between review tone and ratings were flagged and analysed to better understand biases or misinterpretations.

The dataset was further enriched by identifying “Stack of Learning” keywords, allowing us to associate sentiment with specific technologies or courses (e.g., Python, Data Science, Java). This added a layer of semantic relevance to the sentiment trends.

The analysis ensured:

• Reliability across subjective language.

• Coverage of all reviews including non-text entries.

• Usefulness for both learners and institutions to interpret satisfaction metrics clearly.

This well-rounded approach offers an insightful foundation for decision-making, curriculum improvements, and quality assurance in educational services.

**CHAPTER 5**

**DATA VISUALIZATION IN POWER BI**

* Page 1: Sentiment Overview – Review Based

Purpose: Understand sentiment trends based on the review text content alone.

Key Charts:

* Total Reviews & Reviews With No Text: Highlights overall participation and empty reviews.
* Sentiment Category Share (% Positive, Neutral, Negative): Classifies review sentiments using NLP tools (TextBlob).
* Review Sentiment Distribution Over Time: Tracks how review text sentiments change month-by-month.
* Count of Review ID by Sentiment: Shows absolute review counts in each sentiment class.
* Page 2: Sentiment Overview – Rating Based

Purpose: Analyse sentiment based on the numeric star ratings alone.

Key Charts:

* Rating-Based Sentiment Distribution: Categorized using custom logic (e.g., 4–5: Positive, 3: Neutral, <3: Negative).
* Monthly Rating Sentiment Trend: Time series chart showing rating sentiment patterns over months.
* Comparison Between Text-Based & Rating-Based Sentiment: Offers a dual-view for inconsistency spotting.
* Page 3: Rating Sentiment Trend

Purpose: Dive deeper into course-specific sentiment patterns using both text and ratings.

Key Charts:

* Time Series: Rating Sentiment by Month.
* Course Stack Popularity by Review Volume.
* Review Volume by Standardized Course Stack: Shows % split of sentiment types.
* Rating vs Polarity – Sentiment Distribution: Compares textual polarity against actual ratings.
* Technology Mentions by Stack and Sentiment.
* Sentiment-wise Tech Mentions.
* Page 4: Sentiment Deep Dive

Purpose: Explore in-depth sentiment behaviours by course and keywords.

Key Charts:

* Stack-wise Sentiment Breakdown(Rating-Based): Compares Rating Sentiment distribution for each stack.
* Monthly Sentiment Trend: Tracks changes in textual sentiment over time.
* Sentiment Density by Rating and Polarity: Bubble chart illustrating relationship density.
* Course Popularity vs Sentiment Polarity: Highlights course sentiment intensity versus reach.
* Stack-wise Technology Mention Matrix: Matrix of keywords across stacks and months.
* Page 5: Company Sentiment

Purpose: Compare institutions based on sentiment and review volume.

Key Charts:

* Top Institutes by Positive Sentiment Count
* Review Sentiment Distribution by Institution
* Monthly Review Trends by Institution
* Page 6: Review Insights

Purpose: Provide an interactive review-level exploration.

Key Charts:

* All Review Explorer: A table view allowing reviewers, keywords, date, stack, and sentiments to be filtered and searched.
* Sentiment Proportion Based on Review Keywords: Donut chart showing keyword-tagged reviews sentiment split.

**CHAPTER 6**

**KEY INSIGHTS**

* Overall Positivity: Over 85% of reviews—both text-based and rating-based—reflect positive sentiment, suggesting general satisfaction among learners.
* Technology Mentions: Python, Data Science, Java, and Full Stack topped keyword frequency, indicating their high relevance and popularity in Kochi’s Edu Tech offerings.
* Sentiment-Rating Mismatch: Several reviews showed positive ratings but negative or neutral text, hinting at possible rating bias or politeness-driven reviews.
* Stack-wise Analysis:
* MERN and General IT stacks received the most mentions and reviews.
* Niche stacks like Flutter had fewer mentions but surprisingly strong positive sentiment.
* Institutional Trends:
* Luminar Technolab and Techbyheart consistently led in volume and positivity.
* Some institutes with fewer reviews had high polarity but lower star ratings—possibly due to inconsistent delivery.
* Temporal Patterns:
* Peaks in review activity aligned with course completions or batch enrolments.
* Review tone was generally more positive post-completion, especially in placement-driven courses.
* Review Gaps: A noticeable number of reviews lacked text but had high ratings, reducing sentiment clarity.
* Keyword Sentiment: Reviews containing specific tech keywords like “Python” or “JavaScript” tended to be more polarized—either strongly positive or negative—than generic course reviews.

**CHAPTER 7**

**CONCLUSION**

This project effectively captured the real voice of learners by merging structured star ratings and unstructured text reviews into a unified sentiment analysis framework. Through Python-based automation, keyword extraction, and dual sentiment evaluation, it provided a multi-dimensional view of learner feedback in Kochi’s growing Edu Tech sector.

By designing an interactive Power BI dashboard, this project showcased:

* The popularity and reception of specific technologies/courses.
* The reputation trends of top training institutes.
* The presence of rating-text mismatches, offering valuable signals for quality review.

This analysis enables institutes to:

* Tailor their curriculum based on real feedback.
* Benchmark themselves against competitors.
* Track student satisfaction trends over time.

As a future data scientist, this project allowed me to apply:

* End-to-end data pipeline creation
* Real-world NLP
* Automated review extraction
* Dashboard storytelling

**CHAPTER 8**

**FUTURE ENHANCEMENTS**

* Integrate ML models (e.g., Logistic Regression, Random Forest) for predictive sentiment scoring.
* Automate daily or weekly scraping using CI/CD pipelines and cron jobs.
* Expand to other review platforms like JustDial, MouthShut, and video reviews.
* Add Word Cloud visualizations for frequent review themes. Perform topic modelling (LDA) to discover underlying themes across sentiments.