

Leveraging Trustpilot Reviews Using AI and Qualitative Analysis for Strategic Insights at JYSK

By René Gulager Larsen

05/01-2025

Abstract	5
Introduction.....	6
Background	6
Data in JYSK.....	8
Current state in JYSK	8
Challenges	8
Opportunities	8
Conclusion on current state	9
Customer feedback as data	9
Analysis of Trustpilot data.....	9
Quantitative data	10
Qualitative data	10
Conclusion on feedback as data	11
The strategic value of customer feedback	11
Creating a continual feedback loop	12
Benchmark against competitors	12
Cultural and strategic alignment	13
Identifying pain points and being proactive	13
Enhancing decision making and result evaluation.....	13
Conclusion on customer feedback.....	13
Problem statement	14
Central question	14
Objectives:	14
Limitations	14
Project	15
Objectives	15
Target audience	15
Approach description	16
Expected deliverables	16

Methode.....	17
Data collection	17
Data Cleaning.....	17
Summarization	18
Vectorization	18
Storage in vector database	20
Clustering	20
Visualization.....	20
Validation and evaluation	22
Cluster map as a data source for JYSK	22
Limits in this approach	22
The Future of this approach	23
Integration with organizational strategy.....	23
Skills and resources	24
Risks	24
Bias	24
Data quality.....	24
Ethics and data scraping	25
Compliance with Trustpilot's terms of service	25
Privacy and data protection	25
Stakeholder impact.....	25
Lost context	25
Oversimplification of reviews	26
Providing wrong Information to the organization	26
Lack of human quality control.....	26
Conclusion and summary for risks and pitfalls in analyzing Trustpilot reviews for JYSK	26
Conclusion	27

Abstract

This project addresses JYSK's challenge of leveraging unstructured customer feedback from Trustpilot reviews to inform strategic and operational decision-making. The initiative focuses on transforming this feedback into actionable insights by applying a combination of artificial intelligence (AI) techniques and qualitative analysis.

The methodology includes data scraping, cleaning, and summarization using natural language processing (NLP), vectorization, and clustering algorithms such as K-means and DBSCAN. These processes organize reviews into key themes, such as "delivery issues" or "customer satisfaction," enabling the identification of recurring patterns and trends. Interactive visualizations, including cluster maps and sentiment overlays, make these insights accessible to stakeholders across customer service, marketing, and operations.

The findings show the feasibility of combining AI-driven clustering with qualitative methods to prioritize areas for improvement. Ethical considerations, including GDPR compliance, are integrated throughout the project to ensure responsible data usage.

The implications of a proof-of-concept (PoC) project could provide JYSK with a scalable process for integrating customer feedback into decision-making processes.

Introduction

Background

JYSK, a Danish retail company, has grown from its humble beginnings in 1979 to become a global leader in home furnishings and interior products. With operations spanning numerous countries, JYSK continues to expand its presence, bringing Scandinavian design and quality to customers worldwide.

AI has a growing focus in most companies, and JYSK is no exception to that, but until now there have only been small initiatives to start using AI, and the initiatives have mostly been about developers and code automation. For a company like JYSK, customer review platforms such as Trustpilot offer a lot of data that reflects customer experiences, satisfaction levels, and expectations. However, this data is often underutilized due to its volume¹ and unstructured nature. JYSK, like many large-scale organizations, faces challenges in transforming this data into actionable insights. The underutilization of Trustpilot reviews may represent a missed opportunity for JYSK to leverage a source of customer feedback that could lead to strategic and operational improvements. With thousands of reviews collected annually, JYSK has access to data that reflects customer experiences across multiple channels online and in-store, and in various regions. However, without a systematic approach to analyzing and interpreting this data, it remains a largely untapped resource. Understanding how to analyze and utilize Trustpilot reviews systematically presents an opportunity to enhance its service quality, address customer pain points, study competitors, and align organizational efforts with customer expectations. This study explores how AI, clustering, and qualitative interpretation can be employed to leverage Trustpilot reviews effectively while addressing ethical considerations, organizational impacts, and resource requirements.

My interest in this project stems from my daily work, which involves automating tasks for developers to enhance efficiency and streamline processes. Automation, as a concept, becomes even more intriguing when combined with AI to tackle tasks that are typically labor-intensive and time-consuming. One such "trivial" yet significant task is the evaluation of large volumes of customer feedback. By leveraging AI, this project explores how automation can transform raw customer reviews into actionable insights, enabling organizations like JYSK to make data-driven decisions with minimal manual effort. This blend of AI and automation not only addresses scalability but also opens new possibilities for enhancing strategic processes across the organization.

¹ Currently over 31000 reviews on Jysk Damark <https://www.trustpilot.com/review/www.jysk.dk>

By focusing on JYSK's context, this report tries to bridge the gap between raw customer feedback and actionable business insights while highlighting the challenges, possibilities, and future development for using the data. This project seeks to systematically analyze Trustpilot reviews using AI, addressing the current underutilization of this data while aligning with organizational values.

Data in JYSK

JYSK collects thousands of customer reviews on Trustpilot annually, offering a valuable mix of structured star ratings and unstructured customer comments. These reviews span both their online business and retail stores across various countries, providing a diverse dataset reflecting a wide range of customer experiences and expectations. However, this wealth of information is currently underutilized in strategic processes. Properly harnessing this data could enable JYSK to derive actionable insights that enhance decision-making and align with customer needs across markets.

Current state in JYSK

Challenges

Despite the abundance of data, JYSK currently underutilizes it in strategic processes due to several challenges:

- **Volume and Scalability:** The sheer number of reviews makes manual analysis impractical.
- **Unstructured Nature:** Qualitative feedback lacks categorization, making it difficult to identify actionable themes.
- **Data Silos:** Trustpilot data is not integrated with internal systems such as CRM or sales platforms, limiting cross-functional insights.

This is an area that has gained more focus in JYSK and is a part of the strategic decisions going forward.

Opportunities

- By better leveraging the data provided by Trustpilot, JYSK should be able to:
- **Enhance Customer Service:** Identify recurring complaints, such as delivery delays or product quality issues, to proactively address customer pain points.
- **Inform Product Development:** Use qualitative feedback to refine product designs and offerings.
- **Benchmark Performance:** Analyze trends in customer sentiment, measure performance against competitors, and adjust strategies accordingly.

Conclusion on current state

Trustpilot reviews represent an underused resource for JYSK, with the potential to enhance decision-making and align better with customer needs. By addressing current challenges and leveraging the Trustpilot reviews, JYSK can potentially transform unstructured feedback into actionable insights, driving improvements across the organization.

Customer feedback as data

Platforms like Trustpilot have transformed the way organizations access and use customer feedback. These platforms provide companies with an opportunity to gather authentic, real-time insights into customer experiences, satisfaction, and expectations directly from their audience. For JYSK, Trustpilot data offers unique value, as it reflects both quantitative metrics like star ratings and qualitative insights from detailed reviews and responses.

Analysis of Trustpilot data

A typical Trustpilot review provides a wealth of information that can offer valuable insights for JYSK². This dataⁱ includes both **quantitative elements**, such as star ratings, review counts, verification status, and the reviewer's country, and **qualitative aspects**, such as the title, detailed review text, and JYSK's responses. Together, these components provide a comprehensive view of customer sentiment and experiences.

However, this data is not available for direct download, requiring scraping methods to collect and structure it for analysis. Tools like Python's Scrapy can be used to extract the data systematically, ensuring compliance with ethical and legal standards. Once obtained, this data can serve as a rich resource for JYSK, enabling deeper insights into customer feedback. By leveraging this information, JYSK can drive improvements in various areas of the organization, such as refining product offerings, enhancing customer service, and benchmarking performance against competitors.

The combination of structured and unstructured data within Trustpilot reviews makes it an invaluable asset for understanding customer needs and aligning organizational strategies with evolving market demands. This approach allows JYSK to transform underutilized feedback into actionable insights that support data-driven decision-making.

² <https://www.trustpilot.com/review/www.jysk.dk>

Quantitative data

Trustpilot reviews provide a wealth of **quantitative data** that is relevant for JYSK. This data includes measurable elements that can be analyzed. Key quantitative data points include:

Reviewer Name

- *Description:* The name or pseudonym of the person leaving the review.
- *Example:* "Ömer Peterson" (*changed for privacy*)
- *Usage:* Identifies unique reviewers but should be anonymized to ensure privacy.

Reviews Count

- *Description:* The total number of reviews the user has left on Trustpilot.
- *Example:* "1 review"
- *Usage:* Helps assess the credibility of the reviewer (e.g., experienced reviewers vs. one-time reviewers).

Country

- *Description:* The geographic location of the reviewer, if available.
- *Example:* "N/A, DKK, GB etc"
- *Usage:* Enables analysis of geographic trends and cultural differences in customer feedback.

Review Date

- *Description:* The date and time when the review was posted.
- *Example:* "2023-10-26T12:12:45.000Z"
- *Usage:* Useful for trend analysis and monitoring customer sentiment over time.

Star Rating

- *Description:* A numerical score reflecting the customer's overall experience, typically ranging from 1 to 5 stars.
- *Example:* "1 star"
- *Usage:* Enables quantitative analysis of customer satisfaction levels.

Qualitative data

Trustpilot reviews offer a large amount of qualitative data that has potential for analysis. This unstructured feedback provides insights into customer experiences, sentiments, and expectations, key datapoints are:

Review Title

- *Description:* A brief title summarizing the review.
- *Example:* " Great price, great salesperson "
- *Usage:* Highlights key themes or sentiments in the review.

Review Text

- *Description:* The detailed feedback provided by the customer.
- *Example:*
"In the context of customer satisfaction, Victoria, who works at Jysk Glostrup, provided us with a direct exchange for the pillows we had purchased previously. She spent several minutes explaining what was needed for a comfortable night's sleep to both me and my spouse. We are very happy to receive such service. Thank you."
- *Usage:* Offers rich qualitative insights into specific customer experiences, enabling thematic analysis.

Company Reply

- *Description:* A response from the company addressing the review.
- *Example:*
"Hi Ömer, Thank you for your review and the 5-stars ★★★★★. I am so happy to read about your experience with Victoria in Glostrup, and that she exchanged and explained about the pillows. That is the way we want it to be 🙌. I will give the review to Victoria. I'm sure she will be happy about it. I wish you a great evening. //Mette - JYSK Customer Service."
- *Usage:* Provides context on how the company handles feedback, offering insights into customer service strategies.

Conclusion on feedback as data

The data available in Trustpilot reviews offers an interesting possibility for extracting actionable insights for JYSK, with **quantitative data** such as star ratings, review counts, and timestamps and **qualitative data** like detailed review texts and company responses. The dataset provides a direct view of customer experiences and expectations. These data elements enable JYSK to conduct an analysis, from identifying recurring pain points to benchmarking against competitors and monitoring trends over time. The data also shows some problems that are worth considering regarding data privacy, for example, there is a username and the name of a specific person in a store. This should be handled correctly; even though it is data scraped from a public source on the internet.

The strategic value of customer feedback

Customer feedback is a powerful asset for any organization, providing direct insights into customer experiences, expectations, and satisfaction. Reviews not only reflect customer

sentiment but also offer a continuous stream of real-time data. For JYSK, leveraging customer feedback goes beyond resolving individual complaints. It could serve as a foundation for benchmarking against competitors, identifying recurring issues, and proactively adapting to changing market conditions. This section explores how customer feedback can become a strategic resource for continuous improvement and informed decision-making.

Creating a continual feedback loop

Customer feedback serves as a continuous stream of real-time insights, reflecting customer expectations. By leveraging this ongoing feedback, JYSK has the opportunity to monitor shifts in customer sentiment over time, identify emerging trends, and adapt its strategies proactively. For example, a recurring theme in reviews about delivery or product satisfaction can indicate the need for improvements, while positive feedback can highlight strengths to build upon. Establishing a continual feedback loop ensures that customer insights are not only captured but also consistently integrated into decision-making.

Benchmark against competitors

Trustpilot data makes it possible for organizations to compare their performance against competitors. By analyzing customer reviews, JYSK can identify areas where it excels and areas where it falls short compared to industry peers. This comparative analysis provides insights into how JYSK can refine its strategies and operations to maintain a competitive edge. For instance, if competitors receive consistent praise for faster delivery times, JYSK can prioritize logistics improvements to meet customer expectations. Benchmarking not only helps JYSK understand its position in the market but also highlights opportunities for differentiation and growth. An example could be "Consistent praise for competitor X's fast delivery times highlights an opportunity for JYSK to improve its logistics operations." This could help in attaining an actionable insight into the use of benchmarking data to prioritize investments in areas like shipping efficiency or customer service training.

Monitoring competitors using Trustpilot data raises ethical concerns related to transparency and fairness. While analyzing publicly available reviews is legally permissible, organizations should think about how this practice aligns with ethical standards. For instance, using competitor data without acknowledgment could be perceived as

exploitative.

Cultural and strategic alignment

Trustpilot reviews provide insight into whether the organization's values and mission are resonating with customers. This feedback can inform strategic adjustments to ensure alignment with customer expectations, and as an international company, cultures can be different about how customers perceive JYSK

Identifying pain points and being proactive

The data can be used to pinpoint recurring issues in their processes, products, or services for JYSK. For example, frequent mentions of delivery delays can prompt a review of logistics operations. Feedback enables JYSK to proactively address issues before they escalate. This not only improves customer satisfaction but also reduces the likelihood of negative reviews in the future.

Enhancing decision making and result evaluation

By systematically analyzing feedback, organizations can move beyond anecdotal evidence to data-driven decision-making. This helps prioritize actions that align with customer needs, such as refining product offerings, improving customer service, or addressing operational inefficiencies.

Conclusion on customer feedback

Customer feedback is more than just a reflection of customer sentiment. It could be used as a strategic tool for JYSK to help in decision-making, align with market needs, and strengthen its competitive edge. By establishing a continual feedback loop, addressing recurring issues from negative feedback, and leveraging positive feedback, JYSK can use the insights given in the Trustpilot reviews and turn them into actionable insights.

Problem statement

Highlight the gap in systematically leveraging such reviews for actionable insights.

Central question

How can Trustpilot reviews be systematically analyzed to support decision-making processes, improve customer satisfaction, and align with ethical and organizational values?

Objectives:

1. Analyze Trustpilot reviews to identify operational and strategic insights.
2. Explore the ethical and organizational consequences of using AI-driven analysis.
3. Can Ai contribute to giving better insights into what JYSK is of importance to its customers?

These objectives aim to leverage customer feedback as a strategic asset, driving improvements in decision-making, operations, and customer satisfaction while maintaining ethical integrity.

Limitations

For his project I am allowed to use the Trustpilot data from JYSK, it is not possible to use any resources in JYSK, so the views and input are limited from my perspective as an external consultant located on a platform engineering team. This project describes a plan and the thoughts about implementing a solution, it is outside the scope to do the implementation for it.

Project

The project aims to design a scalable process for analyzing Trustpilot reviews using a combination of AI and qualitative methods. This process will help JYSK to extract actionable insights from large volumes of customer feedback, hopefully improving strategic decision-making and customer satisfaction.

Objectives

- **Develop a Proof of Concept (PoC):**
 - Design and test a system for collecting, processing, and analyzing Trustpilot reviews.
 - Focus on scalability, accuracy, and compliance with ethical standards.
- **Extract Actionable Insights:**
 - Identify recurring themes and trends to inform strategic and operational decisions.
- **Demonstrate Value:**
 - Showcase the potential of AI-driven feedback analysis to improve efficiency and customer satisfaction

Target audience

The target audience for this approach encompasses key stakeholders within JYSK who have a vested interest in the proof-of-concept implementation for analyzing Trustpilot reviews. These stakeholders include **Team Management and Executives**, who require actionable insights to align business strategies with customer expectations and drive organizational priorities and are the key to the needed resources for the PoC. The **Customer Service Team** benefits from understanding recurring issues and common pain points, helping them to respond efficiently and enhance the overall customer experience. The **Marketing Team** can leverage insights into customer sentiment to craft targeted campaigns that resonate with specific audience segments, improving engagement and brand loyalty. Future down the line there might be more stakeholders including but not limited to the store manager, legal team, and product development.

Approach description

The implementation involves a structured, step-by-step process designed to transform unstructured customer reviews on Trustpilot into actionable insights. This approach combines AI's computational efficiency with qualitative methods, addressing both scalability and while trying to give a nuanced understanding of customer feedback. Below is a detailed breakdown of the components:

1. Data Collection

- Trustpilot reviews are scraped to extract relevant information, including text, star ratings, and timestamps.

2. Data Cleaning

- Low-quality reviews, such as those from non-verified users or containing insufficient content, are filtered to ensure reliability.

3. Summarization

- Reviews are condensed into concise summaries to streamline the analysis while preserving essential information.

4. Vectorization

- Summarized reviews are converted into numerical vectors that represent their semantic meaning for further analysis.

5. Storage in a Vector Database

- Processed data is stored in a scalable, searchable database for clustering and visualization.

6. Clustering

- Reviews are grouped into clusters based on common themes,

7. Visualization

- Interactive visualizations are created to highlight patterns and insights.

Expected deliverables

Building on the outlined approach, the project aims to produce outputs that investigate the value of analyzing Trustpilot reviews.

- A design of a PoC demonstrating the analysis of Trustpilot reviews for key themes and insights.
- Dashboards or visualizations to present insights dynamically to stakeholders.

Methode

This section provides a detailed explanation of the methods, tools, and techniques used in the process, ensuring that the implementation is scalable, accurate, and ethically compliant. This method focuses on collecting and analyzing the data for JYSK, if there is a need for analyzing competitors the process is close to the same.

Data collection

The data needs to be gathered from Trustpilot, the process for collecting the data is using web scraping.

- **Tool**

Scrapy³, a Python based web scraping framework, is used to crawl Trustpilot and extract reviews.

- **Process**

Reviews are retrieved using scrapy, stored and structured into JSON format, it should capture all data fields such as review text, star ratings, and metadata.

Scrapy has been chosen as it is a well known, and well supported module for python. JSON⁴ is language independent, light data structure that is often used by developers.

Data Cleaning

Data cleaning is needed for a few reasons, to make sure that the data is scraped from Trustpilot is standardized, e.g. if a field is missing a value a new value should be added here to simplify the forthcoming process.

- **Tool:**

Pandas⁵ for data manipulation and regex⁶ for identifying patterns.

- **Process:**

Irrelevant or unreliable reviews are filtered out using a script:

- Exclude reviews from non-verified users.
- Remove duplicates or entries dominated by stop words.
- Remove very short reviews
- Remove personal information

- **Validation:**

Cleaned data undergoes human review to ensure relevance and compliance.

³ <https://scrapy.org/>

⁴ <https://datatracker.ietf.org/doc/html/rfc8259>

⁵ <https://pandas.pydata.org/>

⁶ <https://pypi.org/project/regex/>

The tools chosen are just examples of commonly used tools that could be used for the project, other tools can replace them as this is just string manipulation. The data cleaning process is an ongoing activity and should be improved and verified over time and should be validated by a human review, to verify that no personal information is included, and that the data is correct.

Summarization

Summarizing the data, after the data has been cleaned, is a process where the review itself is condensed down to extract the key elements and there, and this bridges the gap between quality and quotative data, by making the reviews structured into a uniform format, thereby making it possible to analyze.

- **Tool**
OpenAI api⁷ using Python or another programming language
- **Process**
Reviews are condensed into 10-word summaries that capture the essence of the feedback. By posting the review body to the OpenAI Api as a Json object and received as a Json object.
- **Validation**
Randomized spot-checking ensures the summaries retain key customer sentiments.

This part of the method should be tested and reviewed as there are some assumptions about what to do. The choice of 10-word summarization should only be considered a starting point for this PoC, and there should be conducted a test with different lengths. The same is relevant for the choice of using OpenAI's models that OpenAI is offering as there are new models being released often, and the choice of model could influence the quality and the price of the result.

Vectorization

To analyze unstructured customer reviews, the texts can be converted into numerical representations(vector) that AI models can process. This step helps in identifying patterns and grouping reviews into clusters later in the process.

- **Tool:** OpenAI's text-embedding-3-smal or text-embedding-3-large⁸

⁷ <https://platform.openai.com/docs/api-reference/introduction>

⁸ <https://platform.openai.com/docs/guides/embeddings>

- **Process:** Summarized reviews are converted into high-dimensional numerical vectors that represent their semantic meaning.

The tools chosen here should again be tested and evaluated.

Storage in vector database

This is simply to store the data

- **Tool:** Pinecone⁹, a cloud native vector database or pgvectorⁱⁱ.
- **Process:** Vectors and their associated metadata are indexed for efficient retrieval and future scalability.

Clustering

Clustering is grouping reviews with similar characteristics together, making the identification of actionable patterns within large datasets possible. This step helps “sort” unstructured customer reviews into groups that JYSK can then be used to visually identify patterns.

- **Clustering Algorithms:** K-means¹⁰ for scalability and DBSCAN¹¹ for identifying anomalies and outliers. Or similar algorithms.
- **Dimensionality Reduction:** t-SNE¹² or UMAP¹³ for reducing high dimensional data to a 2D or 3D space for visualization.

Clustering gives the possibility to find related reviews and thereby see groups and while combined with a dimensional reduction it is possible to create a visualization of the groups. Choosing the right algorithms for this is outside the scope of the project.

Visualization

Visualization transforms complex clustering data into an intuitive, user-friendly format, allowing JYSK teams to quickly and easily gain insights into the data.

- **Visualization Tools:** Plotly¹⁴ and D3.js¹⁵ are used to create dynamic dashboards and cluster maps.

The tools are chosen as they are already known and supported insight JYSK and should help in simplifying the interpretation of all the data. This should be turned into a live dashboard, so reviews can be monitored in real-time.

⁹ <https://www.pinecone.io/>

¹⁰ <https://www.ibm.com/think/topics/k-means-clustering>

¹¹ <https://en.wikipedia.org/wiki/DBSCAN>

¹² https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding

¹³ <https://umap-learn.readthedocs.io/en/latest/>

¹⁴ <https://santilland.github.io/plotly/index.html>

¹⁵ <https://d3js.org/>

Validation and evaluation

Validation and evaluation are important steps to ensuring the accuracy, reliability, and relevance of the result of the clustering. These processes help refine methodologies and maintain stakeholder relevance of the data.

- **Human Oversight:** Cross-departmental teams review clusters to ensure themes align with actual customer feedback.
- **Evaluation Metrics:** Accuracy of clustering, stakeholder usability of insights, and impact on decision-making processes.

Validation and evaluation are important to ensuring that the results of clustering are accurate, reliable, and relevant. These processes not only refine methodologies but also ensure that the insights generated align with actual customer feedback and stakeholder needs.

Cluster map as a data source for JYSK

A cluster map can serve as a data source by transforming large, unstructured datasets into organized overviews by grouping similar data points based on shared characteristics, it can be used to reveal patterns and trends that might otherwise remain hidden. For example, Trustpilot reviews for JYSK could be grouped into clusters such as "delivery issues," or "customer service satisfaction," making it easier to identify recurring themes and prioritize actions. Anomalies or outliers within the map can also highlight unique feedback or emerging topics, such as niche product issues or isolated complaints. A cluster map can be used as a dynamic tool. This allows for continuous updates and interactive exploration, enabling stakeholders to zoom into specific clusters and analyze individual reviews for deeper context. This approach lowers data complexity, could support decision-makers, and provides a way to monitor trends or the result of strategy changes over time.

Limits in this approach

There are a lot of different areas of the data that could be interesting to investigate and can help in understanding, qualifying, and prioritizing the data thereby affecting how and what the cluster map is presenting. In the process proposed there is just a summarizing of the review by instructing an NLP and limiting it to 10 words. Other things could be interesting to look into, like whether there is a difference in what people rate (stars) and what the sentiment is in the text, *“Is there a difference in countries?” etc.* The proposed method is considered a PoC, that should pave the road for more exploration and learning for the origination.

As a PoC, this method is designed to validate the feasibility of combining AI and qualitative methods, helping the way for more comprehensive analyses, such as multi-country comparisons or sentiment analysis across product categories.

The Future of this approach

The proposed approach, while initially a PoC, has the potential for expansion and evolution, paving the way for more sophisticated and comprehensive analyses in the future. As JYSK gains experience with this process, the framework can be enhanced to accommodate additional data sources, such as reviews from social media platforms or customer surveys, providing a more holistic understanding of customer sentiments. Multi-language support and cross-regional sentiment comparisons can further help with insights, making them actionable across JYSK's markets. Expanding the granularity of analysis, such as comparing sentiment by product categories or customer demographics, can provide deeper insights into specific business areas. The approach also has the potential to be used in different organizational functions, including marketing optimization, product design, and supply chain management.

Integration with organizational strategy

The proposed process fits into the organizational strategy, emphasizing customer-centric decision-making. By transforming customer feedback into actionable insights, this approach helps strategic priorities by organizing data on customer sentiments and experiences. Clustering reviews into key themes could help JYSK to focus on areas of high impact, aligning with the company's goal of improving customer satisfaction.

Furthermore, this process supports JYSK' in getting their "feet wet" introducing AI and data-driven methods into the operational workflow. Insights derived from this process can help in resource allocation and refine marketing campaigns. Additionally, the dynamic nature of the cluster map provides ongoing feedback, enabling JYSK to monitor the success of strategic initiatives and adapt quickly to changing customer needs. When the PoC is done and if it has matured a bit, there should be introduced some KPI¹⁴, to measure the applied usability of it, a KPI could be an improvement of the general score on Trustpilot.

Skills and resources

To start the project, JYSK requires a combination of technical skills, organization, and resources. Key technical skills include expertise in data scraping, cleaning, and managing vector databases, as well as some kind of understanding of natural language processing (NLP) and clustering algorithms for analysis and visualization. The amount of organizational resources such as scalable infrastructure and visualization tools, are quite limited and have next to no cost. However, it is important for JYSK to seek expert advice on GDPR compliance to ensure that data collection and processing adhere to all relevant privacy regulations, particularly given the potential for handling customer reviews containing personally identifiable information.

Risks

While the integration of AI and qualitative methods for analyzing the reviews offers numerous advantages, it also brings inherent risks and biases. These challenges need to be understood and mitigated to ensure the process generates meaningful and ethical insights.

Bias

Analyzing Trustpilot reviews for JYSK could distort insights and lead to wrong decisions. Potentially polarized feedback from customers, differences in demographics, or regional differences for the groups of customers using Trustpilot can impact the insights gained from the analysis. Revive feedback from stakeholders and should be used for validating the result.

Data quality

Data quality is a critical factor in ensuring that insights derived from Trustpilot reviews are accurate, reliable, and actionable for JYSK. Poor data quality can lead to misinterpretations, biased conclusions, and ultimately misguided decisions.

Without high-quality data, clustering and summarization algorithms may produce misleading patterns, and subsequent insights could result in inappropriate organizational decisions. For example, inconsistent or biased data could lead JYSK to prioritize solving the wrong customer issues or invest in areas that do not yield significant improvements. To ensure the data quality, there should be an effort to validate, clean, handle empty fields, and logging of errors in the data.

Ethics and data scraping

Data scraping raises several ethical concerns that must be addressed to ensure compliance with legal standards and respect for the rights of individuals and organizations. When scraping Trustpilot reviews for JYSK, the following ethical considerations should be considered.

Compliance with Trustpilot's terms of service

Trustpilot most likely has a TOS, that might limit what is allowed to do with the review data, that might now allow usage of the data or limit it to only use data about JYSK and not competitors. There might be a need to contact Trustpilot to get explicit permission to use the data.

Privacy and data protection

Trustpilot may inadvertently contain personally identifiable information or sensitive data, posing privacy risks. This information could appear in usernames, especially if the same username is used across platforms like Facebook, or within the content of the reviews themselves, where users may include names, contact information, or transaction-specific details. Additionally, metadata such as timestamps or geographic data could potentially identify individuals when combined with other datasets. JYSK should implement data anonymization techniques, such as pseudonymizing usernames, and practice data minimization by collecting only essential fields.

Stakeholder impact

Stakeholders at JYSK, including different departments and management, can be impacted by decisions based on Trustpilot review data, poorly processed or biased data can lead to incorrect insights, which may result in wrong decisions that negatively affect these groups or waste of money, and have an impact on JYSK's competitiveness.

Lost context

Clustering reviews often involves reducing rich, detailed customer feedback into numerical representations, which can strip away the nuance and context of individual reviews.

Clustering reviews in this context involves transforming rich, detailed customer feedback into numerical representations for analysis, which can result in the loss of nuance and context. This loss of context can lead to oversimplified insights, misrepresentation of customer concerns, and ultimately, misguided decisions.

Oversimplification of reviews

Summarizing or categorizing reviews might ignore complex customer sentiments, fail to capture multi-faceted feedback that can have different meanings depending on the country or miss the use of sarcasm.

Providing wrong Information to the organization

Errors in clustering or analysis, such as misclassification or bias in the data, can lead to misleading insights. If these insights are used to inform decisions, JYSK could invest resources in solving non-existent problems or neglecting critical issues. Clustering might overemphasize complaints about a specific product/area due to skewed data, leading management to focus on those products/areas while ignoring broader challenges.

Lack of human quality control

Without sufficient human oversight, AI outputs can go unchecked, amplifying errors or biases in the data. Automated systems might miss subtle cues like sarcasm or cultural references, leading to flawed interpretations. Human quality control is essential to validate the findings and ensure they align with organizational goals and customer realities.

Conclusion and summary for risks and pitfalls in analyzing Trustpilot reviews for JYSK

The integration of AI and qualitative methods to analyze Trustpilot reviews presents both opportunities and challenges for JYSK. While the process can provide valuable insights, it also introduces risks related to bias, data quality, ethical concerns, stakeholder impact, and limitations in data handling. Bias is a central concern, as it can affect insights and lead to unethical or wrong decisions. Similarly, data quality plays a critical role in ensuring insights are accurate and actionable, as poor-quality data can result in organizational decisions that are not based on reality. The main cure for these concerns is human validation, using the stakeholder's feedback and sampling the data to validate the correctness.

Conclusion

This project addresses a gap in JYSK's ability to leverage Trustpilot reviews for actionable insights, as identified in the problem statement. By integrating AI and qualitative methods, it develops a process to transform unstructured feedback into meaningful insights that can inform strategic and operational decisions.

The proposed approach enables JYSK to extract value from its large number of customer reviews, turning the data into a strategic asset. Through clustering and visualization, the project provides tools to identify recurring themes such as delivery issues or customer satisfaction, helping decision-makers to prioritize areas. This data-driven approach can be used to create continuous improvement in customer service, product development, and marketing strategies.

Furthermore, the focus on ethical considerations, including GDPR compliance and risks, ensures that the project aligns with organizational values. The inclusion of feedback loops and stakeholder engagement ensures that the insights generated are not only actionable but also relevant and verified.

As a proof of concept, this project lays the foundation for future scalability, with potential expansions into multi-language support, and cross-regional analysis. By addressing current limitations such as data quality, lost context, and bias, the process can evolve to provide even greater value.

In conclusion, this initiative transforms the underutilized potential of Trustpilot reviews into value for JYSK's customer-centric strategy, aligning data insights with business goals and setting the stage for long-term and competitive advantage.

```
{
  "Reviewer Name": "Prem REDACTED",
  "Reviews Count": "2 reviews",
  "Country": "N/A",
  "Review Date": "2023-07-17T19:22:10.000Z",
  "Star Rating": 1,
  "Review Title": "Great price, great salesperson",
  "Review Text": "I moved into a new apartment and needed to buy lot of the things. Jysk is the best place to find all the furniture and household items with authentic Scandinavian designs. And I was lucky to find a great salesperson, his name is Rune, who helped me to get amazing discounts and he was so friendly that he made whole stressful shopping experience a breeze 😊",
  "Company Reply": "Reply from JYSK Danmark - 2023-07-18T06:19:01.000Z: Hi PremHow wonderful it is to hear about this great experience you have had in connection with your purchase in JYSK.We hope you will be really happy with the new items, and they will give the new apartment the right feeling of home.We are very
```

happy with the 5 nice stars and I will pass them on to Rune, he fully deserves them. ★★★★★ Wish you a really nice day//Maria. JYSK Customer Service"

}

ⁱⁱ <https://github.com/pgvector/pgvector>