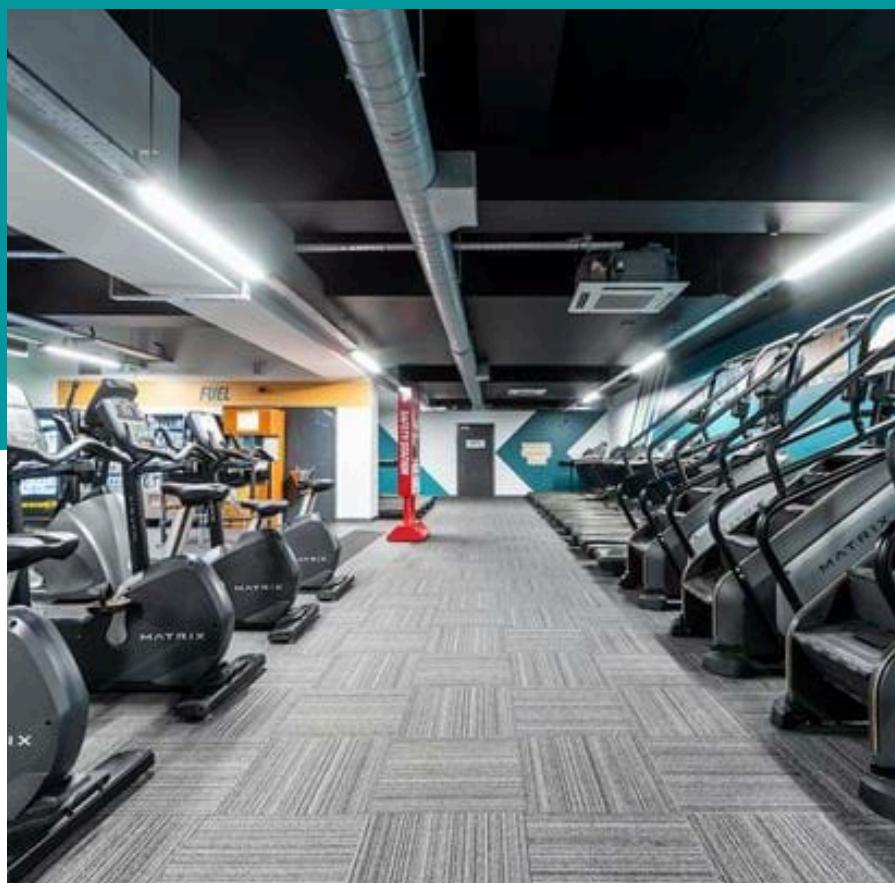


YEAR 2025

# APPLYING NLP FOR TOPIC MODELLING AT PUREGYM



**PREPARED BY:**  
**ONYEKA MUOKA**





# TABEL OF CONTENT

Introduction -----	1
Problem Statement -----	2
Data Description -----	3
Methodology-----	4
Word Frequency and Cloud Analysis (NLTK)-----	5
BRETopic Analysis -----	7
Emotion Analysis (Angry Reviews)-----	10
LLM Analysis of Negative Reviews: Generating Actionable Insights -----	12
Genism LDA Model: Negative Reviews -----	14
Conclusion and Recommendation -----	15

---

# INTRODUCTION

It is vital for businesses to know which areas need improvement to maintain customer experience and retention. PureGym, with 600 clubs and nearly 2 million members, collects extensive customer feedback through online reviews, but the volume makes manual analysis unfeasible. This project uses NLP on Google and Trustpilot reviews to automate topic modelling, aiming to identify factors driving negative feedback to help PureGym improve member experience.

BERTopic and Gensim's LDA is applied to uncover recurring complaints, operational issues, and actionable insights across locations. Word frequency checks, wordclouds, and emotion analysis help isolate strongly negative reviews. We also explore large language models like Phi-4-mini-instruct to summarise key topics and suggest improvements. Due to computing limits, the project focuses only on review data, demonstrating how automated topic modelling can help PureGym address member concerns while delivering affordable, quality fitness services.

---

# PROBLEM STATEMENT

Although PureGym gathers a lot of client reviews, it doesn't have a methodical approach to figuring out what causes unfavourable comments. This hinders its capacity to promptly resolve problems, increasing the possibility of a bad customer experience and member churn. In order to enable focused enhancements to the member experience, this project intends to use NLP-based topic modelling to identify the primary causes of unfavorable evaluations.

---

# DATA DESCRIPTION

The project makes use of consumer review data from Trustpilot and Google Reviews, which includes comments from PureGym members in various locations over a 12-month period. There are roughly 23,250 records in the Google dataset and 16,673 rows in the Trustpilot dataset. The review content, date, rating, and location are all included in each submission.

A number of preparation procedures were used to guarantee consistency and quality before the data was ready for analysis. To cut down on noise, columns that had no analytical value; such as anonymous user IDs, survey IDs, URLs, and reply dates, were eliminated. To keep text analysis consistent, reviews written in languages other than English were also disregarded. To concentrate primarily on insightful comments, any rows with null or empty review content were removed. To increase the accuracy of the NLP models, the review text was standardized by converting it to lowercase, eliminating punctuation, and eliminating excessive whitespace.

---

# METHODOLOGY

Google and Trustpilot reviews were imported, cleaned, and filtered for English-only content. Reviews were tokenized, and stopwords, punctuation, and numbers were removed using NLTK.

Top 30 shared locations were identified, and reviews with ratings below 3 were extracted for negative sentiment analysis. Word frequencies were visualized to spot recurring themes.

BERTopic was applied to these negative reviews to extract key topics, supported by heatmaps and word bar charts for exploration. A BERT-based emotion model isolated anger-heavy reviews, which were reprocessed through BERTopic to uncover precise drivers.

Phi-4-mini-instruct was used to summarize topics and generate recommendations. Gensim's LDA was also applied for benchmarking, allowing a structured comparison with BERTopic.

This layered NLP workflow enabled systematic extraction of actionable insights to address member concerns at scale.

# WORD FREQUENCY AND CLOUD ANALYSIS (NLTK)

## All common locations in both dataset all reviews

Strong positive sentiment was reflected in both the Google Reviews and Trustpilot charts and word clouds, with shared keywords such as equipment, great, good, staff, clean, friendly, machines, always, really, and classes indicating overall positive perceptions of the company. To enable deeper insight extraction, the analysis was subsequently focused on the negative reviews.

## Negative Reviews from Shared Locations

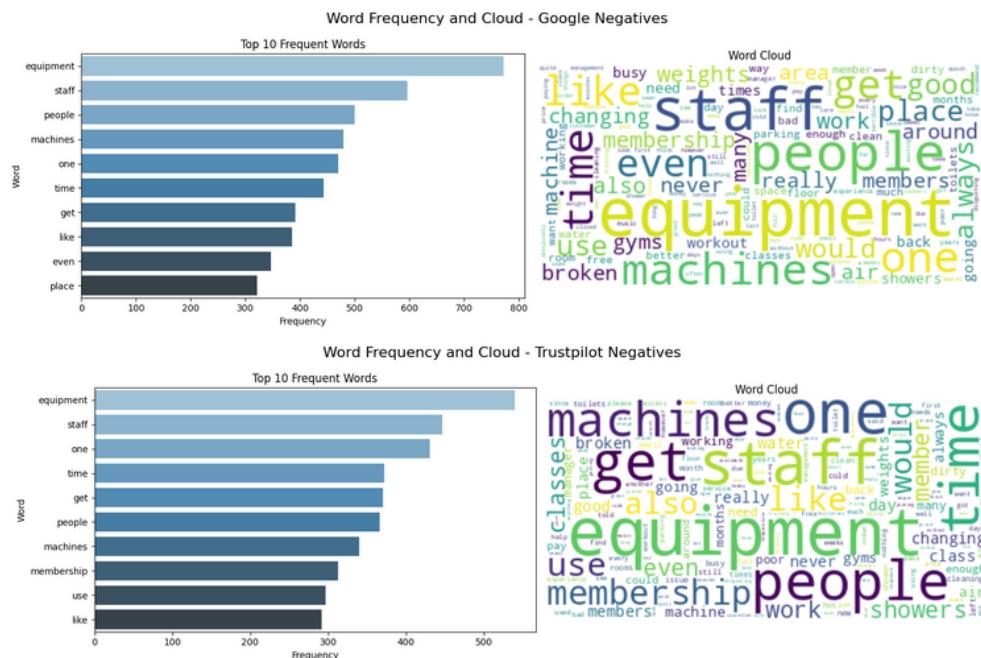


Figure 1. Word frequency and cloud for negative Reviews from Shared Locations

When narrowed to negative reviews in common locations, frequent words indicated concerns with equipment, staff, membership, and location. Repeated mentions of equipment, staff, and machines suggested maintenance and service issues, while terms like one, time, and get pointed to access difficulties. References to people, place, and membership reflected frustrations with crowding, environment, and billing.

## Top 30 Common Locations: Negative Reviews

Compared to the initial run of the negative reviews on all common locations, similar themes were observed, with slight shifts in word frequency. Terms such as equipment, people, staff, machines, one, time, even, always, get, and like remained consistent, though “people” and “always” appeared more prominently. This suggested that complaints about crowding and recurring issues that were being reported frequently by users.

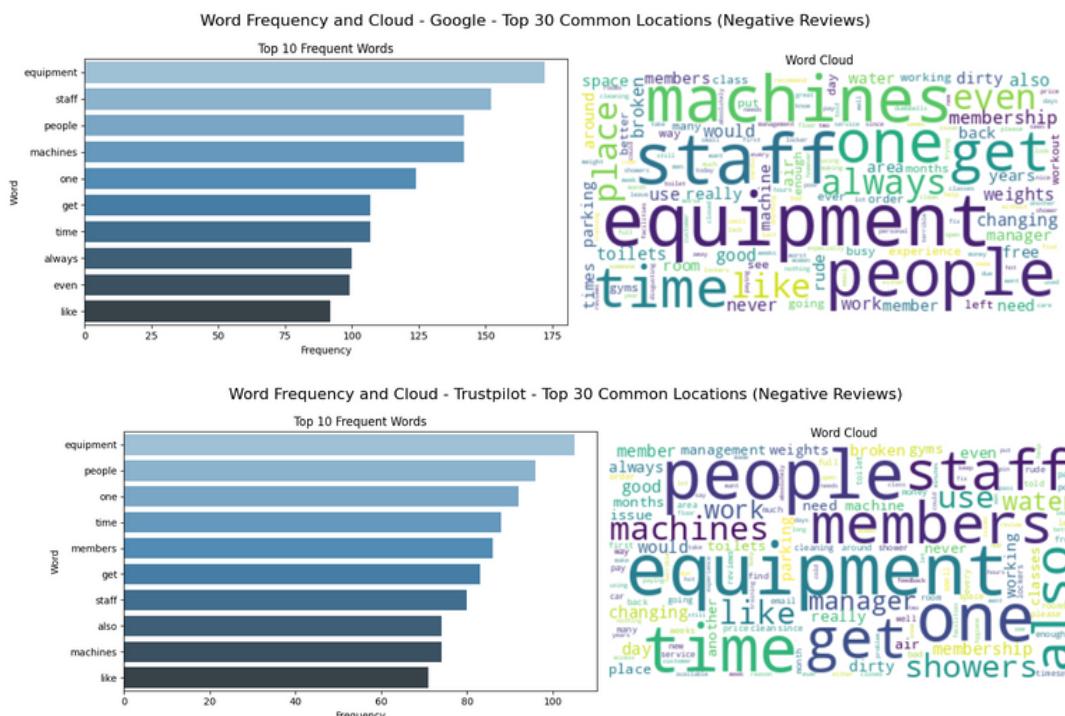


Figure 2. Word frequency and cloud for negative Reviews from top 30 common location

# BERTOPIC ANALYSIS

BERTopic was used to further extract and visualize key topics clusters from cleaned negative reviews

## Negative Reviews from Shared Locations

BERTopic analysis revealed 50 topics from negative reviews in common locations, with the top 10 visualized below.

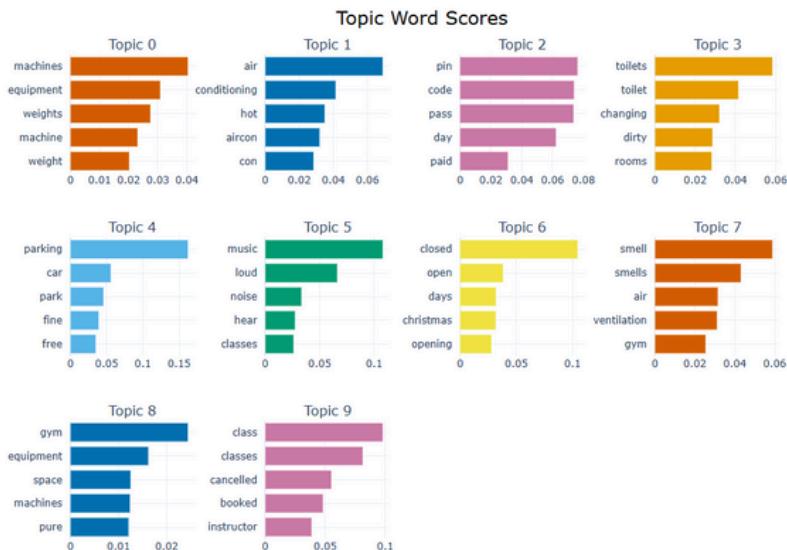


Figure 3. BERTopic Top 10 topics; Negative Reviews from Shared Locations

The identified topics were grouped into related thematic clusters. Below, we present and describe 10 of these key clusters.

Cluster	Theme	Keywords	Description
1	Shower/ Temperature Issues	showers, cold, water, aircon, hot, temperature, freezing, dirty	Complaints about cold or broken showers, poor water pressure, and malfunctioning air conditioning that cause discomfort during workouts and after training.
2	Membership & Billing Problems	membership, cancel, fee, charged, discount, joining, paid	Frustrations with hidden fees, membership cancellations, and issues with obtaining promised discounts.

Cluster	Theme	Keywords	Description
3	Equipment Overcrowding &	machines, broken, wait, overcrowded, busy, equipment, order	Complaints about cold or broken showers, poor water pressure, and malfunctioning air conditioning that cause discomfort during workouts and after training.
4	Facility Cleanliness & Maintenance	cleanliness, toilets, dirty, showers, changing	Reports of unclean toilets and showers, poor waste management, and inadequate upkeep of shared spaces.
5	Hygiene Complaints	smell, dirty, cleaning, toilets, bins, sweat, hygiene	Complaints about unpleasant smells, dirty toilets, overflowing bins, and lack of cleaning supplies like wipes and sprays.
6	Noise & Conduct Issues	music, loud, kids, phones, rude, genocide, boycott	Issues with loud music, disruptive behavior, and broader controversies linked to the company
7	Locker & Security Problems	lockers, stolen, bag, padlock, room, space	Reports of locker theft, damage to personal belongings, and insecure locker setups that reduce trust in the facility.
8	Equipment Availability & Layout	equipment, small, busy, weights, cable, wait, crowded	Challenges with equipment availability, cramped layouts, crowded free-weight areas, and poorly organized spaces.
9	Weight Area Organization	plates, weights, bench, bars, squat, safety	Reports of misplaced plates and bars, crowded squat racks, and poorly maintained benches that create safety concerns.
10	Access & Operational Policies	closed, open, parking, fine, code, pass, induction, door	Dissatisfaction with limited opening hours, inconvenient inductions, parking fines, and entry system issues.

Table 1. BERTopic Clusters; Negative Reviews from Shared Locations

## Top 30 Common Locations: Negative Reviews

Alternatively we ran BRETopic for negative reviews from the top 30 common locations. This led to more precise and actionable subtopics that highlights customer frustrations.

- **Air conditioning** complaints divided into separate issues of gyms being too hot versus too cold, while shower-related complaints split into cold water, fungal growth, and dirty conditions.
- **Equipment** complaints became more detailed, revealing issues with specific machine types, refurbishment delays, and broken cables.
- **Membership** issues separated into concerns about fees, subscription problems, joining discounts, and suspension processes.
- **Location-specific mentions** surfaced, with gyms in places like Bradford, Finchley, Abbey Wood, and Tottenham Court named directly in complaints.
- **Staff rudeness** with specific names mentioned, and concerns around voice or mental bullying within gyms.

Other topics came up including issues with kids around the gym area, wifi connectivity, fire doors.

Narrowing down this run highlighted some precise topics compared to the first analysis.

# EMOTION ANALYSIS (ANGRY REVIEWS)

To dig deeper, we imported the bhadresh-savani/bert-base-uncased-emotion model from Hugging Face to identify reviews where anger was the top emotion. We then isolated these reviews for a focused BERTopic analysis.

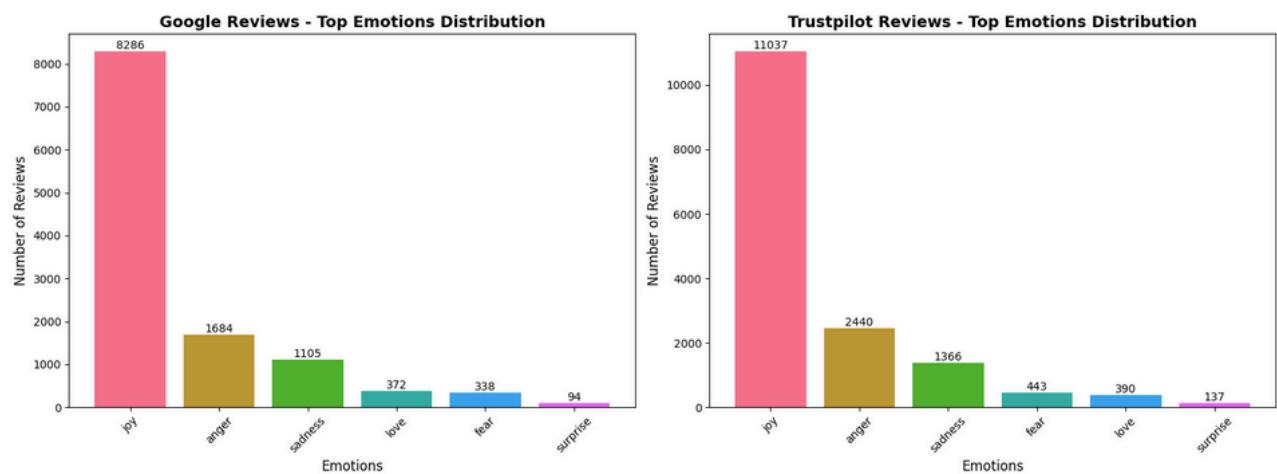


Figure 4. BERT analysis: Distribution of Top Emotions in Reviews from Shared Locations

After analysing these anger-tagged reviews with BERTopic, the top 10 themes remained consistent, but the order shifted.

Air conditioning issues moved to the top, showing it as a key driver of angry reviews.

Toilet problems, staff rudeness, and gym openings on Christmas Day also moved higher, indicating these issues trigger strong negative reactions from customers.

Other topics emerged outside the top 10, including mentions of Yanga water (sports drink), rowing machines, and squat racks, reflecting additional sources of frustration, though less frequent.

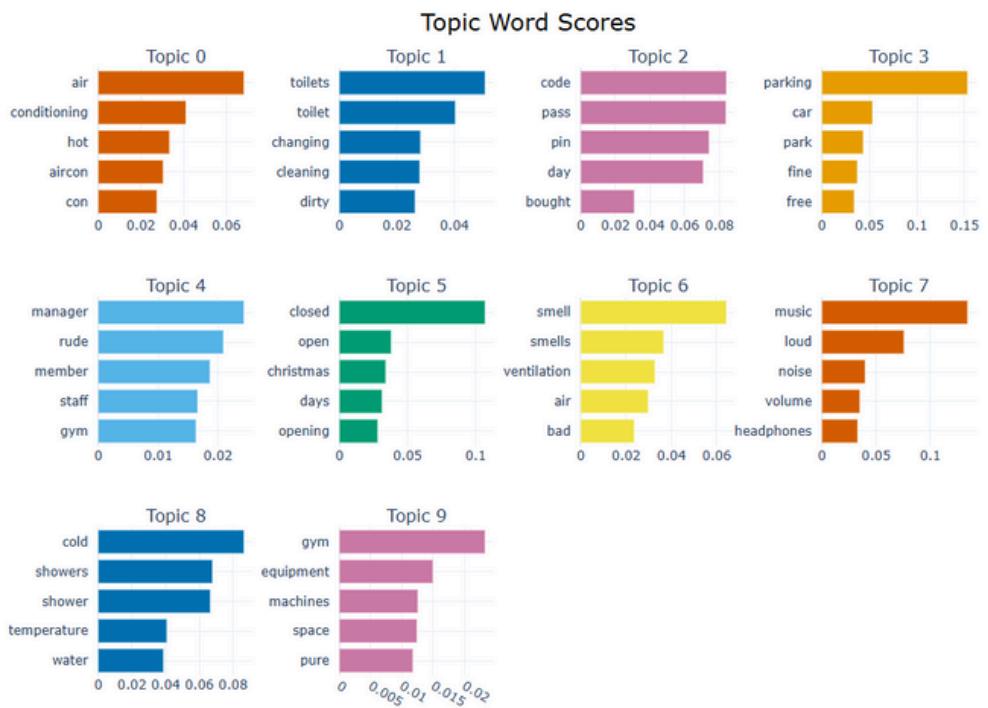


Figure 5. BERTopic Top 10 topics; Angry Reviews

# LLM ANALYSIS OF NEGATIVE REVIEWS: GENERATING ACTIONABLE INSIGHTS

## Complaints topic

The Phi-4-mini-instruct model was used to summarise key topics from angry reviews within the top 30 common locations. These summarised topics were then passed into BERTopic to enable further refinement and narrowing down of the key issues raised.



Figure 6. BERTopic Top 10 topics; LLM output

Recurring themes included air conditioning, ventilation, machine and equipment repairs, parking, and hygiene facilities. The highest-ranking topic, however, was membership cancellation and communication support, highlighting these as another main drivers of negative reviews.

Additional topics emerged, though outside the top 10, including mirror space, privacy concerns, and complaints about the treadmill.

## **Insights Generated.**

The model was also used to generate suggestions which are noted in the table below

Area	Actionable Insight
Facility Comfort and Air Quality	Invest in high-quality air conditioning and ensure regular maintenance to improve comfort.
Cleanliness and Maintenance	Implement a rigorous cleaning schedule and maintain all high-traffic areas consistently.
Availability and Accessibility	Improve booking systems, consider extended hours, and add more classes to manage demand.
Parking and Signage	Enhance signage with clear, well-lit instructions and explore adding parking spaces or partnerships.
Equipment and Service Quality	Regularly update and maintain equipment, and train staff to deliver attentive service.
WiFi and Entertainment	Ensure reliable, fast WiFi and consider more music options or a dedicated entertainment area.
Regulation and Safety	Install clear safety signage and maintain availability and quality of safety equipment.

*Table 2. Insight Generated; Phi-4-mini-instruct model*

# GENISM LDA MODEL : NEGATIVE REVIEWS

For comparison Genism LDA model was employed to the negative reviews from all the common locations. the top 10 topics are shown below

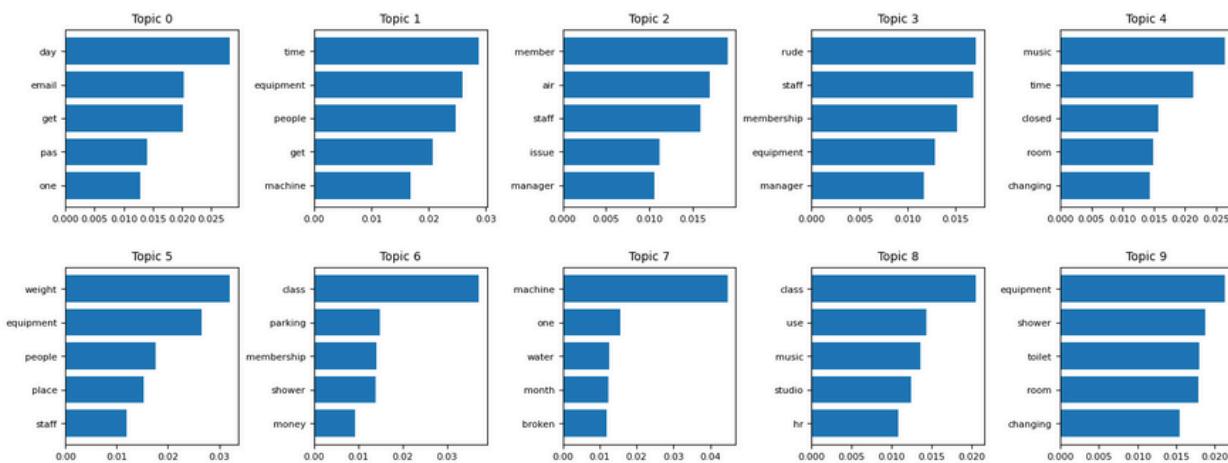


Figure 7. Genism LDA Model: Top 10 topics

The LDA model identified similar themes as the BERTopic model, including equipment and machines, instructor classes, management and staff, hygiene, membership, and air conditioning.

However, when extracting keywords for each topic, LDA sometimes included unrelated words. This may reflect limitations in fine-tuning. For instance, Topic 3 highlighted “member,” “air,” “staff,” “issue,” and “manager” as top keywords, even though they are unrelated. Similarly, the word “equipment” appeared frequently in many topics, reducing clarity in topic separation.

---

# CONCLUSION AND RECOMMENDATIONS

## Conclusion

Recurring concerns among PureGym users have been identified using BERTopic, LDA, and emotion models. These concerns, surfaced across key locations, include:

- Air conditioning and ventilation
- Equipment faults and repair delays
- Hygiene in showers, toilets, and changing rooms
- Membership cancellations and poor communication
- Staff and management rudeness
- Parking facilities
- Overcrowding

## Recommendation

A drill-down analysis is recommended to clarify the specific issues within each topic. Dashboards should be created to track these topics over time, allowing improvements to be measured.

Additional compute resources should be allocated to enable the analysis of non-English reviews, ensuring all member feedback is captured. Engagement with an industry expert is suggested to validate these insights and align strategies with practical needs.