Project Report

Lead Persona Enrichment Tool

-(Using Generative Large Language Models for B2B Sales Outreach)

Developed by

Name - Dhairya Kikani
Phone No. - +91 8780569713
Email - kikanidhairya@gmail.com
Github - https://github.com/dhairya-kikani
LinkedIn - www.linkedin.com/in/dhairya-kikani-5b1b97214

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to the mentors, reviewers, and organizers who provided the opportunity to work on this project. Special thanks to the developers of Groq and Streamlit for creating powerful tools that enabled seamless integration of AI and user interface components.

I also extend my appreciation to the open-source community for continuous contributions that simplify the development of real-world applications using large language models (LLMs). Lastly, I acknowledge the support and feedback from peers and colleagues who offered valuable insights during the development of this tool.

This project was a great learning experience in merging large language models with business processes, and I am thankful to have had the opportunity to explore it thoroughly.

ABSTRACT

In modern B2B outreach, personalization is key to capturing the attention of potential clients. Sales professionals often work with raw lead data extracted from LinkedIn, lacking the context necessary for meaningful engagement. This project presents a Lead Persona Enrichment Tool that leverages large language models (LLMs) via the Groq API to generate enriched lead profiles with three tailored insights: a persona summary, a personalized pitch angle, and a contextual icebreaker.

The tool automates what would otherwise be a manual and time-consuming process, significantly improving both efficiency and quality of outreach. The enriched data is presented through a Streamlit dashboard with dynamic filtering and export functionality. This approach not only saves time for Sales Development Representatives (SDRs) but also aligns closely with business objectives by enabling more relevant and effective lead communication.

Table of Contents

1. INTRODUCTION

Modern B2B sales teams face challenges when attempting to personalize outreach at scale. Sales Development Representatives (SDRs) often work with raw LinkedIn lead exports containing limited information — typically just name, job title, and company. This results in generic outreach that fails to capture the lead's interest. To solve this, I have built a Lead Persona Enrichment Tool that uses large language models (LLMs) to generate personalized insights for each lead, boosting outreach relevance and efficiency.

2. Problem Statement

The lack of context around LinkedIn leads forces teams to either spend time researching each profile manually or send impersonal messages. Both approaches are inefficient. My objective was to automate this process by using LLMs to generate:

- A concise persona summary
- A relevant pitch angle
- A contextual icebreaker

3. Objective

- Automatically generate relevant buyer personas for each lead.
- Provide personalized pitch angles for sales outreach.
- Create human-sounding icebreakers to improve engagement.
- Deliver all outputs in a searchable, downloadable interface.
- Make the system scalable, modular, and easy to integrate into sales workflows.

4. Business Use Case Understanding

Sales and marketing teams want deeper insights into potential customers without manual research. This tool is built to:

- Prioritize high-impact leads.
- Reduce research time.
- Align sales messaging with the target customer's mindset.
- Fit smoothly into existing CRM/export processes.

5. System Architecture & Workflow

i.) Input CSV (leads.csv) containing basic LinkedIn data.

ii.) Each row is processed by:

- generate_persona()
- generate_pitch_angle()
- generate_icebreaker()

iii.)Responses are collected and saved in "enriched_leads.csv"

iv.)A Streamlit dashboard visualizes the results with filtering, search, and download options.

6. Model Selection

For this project, I selected the Mixtral-8x7B-Instruct model via the "Groq API". This model was chosen due to its strong instruction-following capabilities, fast inference speed, and ability to generate structured and human-like responses — all crucial for persona, pitch, and icebreaker generation.

7.Data Collection & Preprocessing

- Input format: CSV with fields like Name, Job Title, Company, Industry, LinkedIn
- Pre-checks for missing fields or formatting issues.
- Rows are validated before processing.
- Option to retry failed rows in the future (via try-except block)

8. Prompt Engineering

Each function uses custom prompts based on the lead's details:

- Persona: Summarize likely responsibilities and mindset
- Pitch Angle: Reflect common pain points and value alignment
- Icebreaker: Write a warm, contextual opener referencing the lead's role/company

Prompts are modular, so they can be tested and tuned independently.

9. Implementation Details

- Built in Python
- Uses groq SDK for model calls
- Modular structure:

    groq_prompts.py → LLM prompts

    process_csv.py → CSV enrichment logic

    test_prompt.py → Manual lead testing

    app.py → Streamlit frontend
- Output saved in enriched_leads.csv

10. Streamlit UI Design

- Sidebar filters (Name, Industry, Company)
- Dynamic table for all leads
- Preview panel for detailed view of the first filtered lead
- CSV download button
- Clean UI with minimal setup required

11. Performance Evaluation

i.) Subjective Evaluation:

Persona relevance: ✓

Pitch insightfulness: ✓

Icebreaker tone: ✓

ii.) Enriched leads demonstrate:

Improved personalization in outreach

Clear buyer-focused messaging

iii.) Runtime: Enriches 50–100 leads within a few seconds using Groq's API.

12. Key Features & Benefits

a.) Fast LLM inference via Groq Accurate lead understanding using tailored propmt.

b.) Actionable outputs for sales teams

c.) UI for filtering/exporting

d.) Modular code for easy updates or integrations

13. Limitations

-> Currently relies on manual CSV input
-> Limited error retry logic (no auto-retry mechanism)
-> May need rate limiting for bulk processing
-> Output quality is still based on LLM interpretation (not verified facts.)

14.) Future Enhancement

- CRM integration (e.g., HubSpot/Salesforce)
- Lead scoring or priority tagging
- UI improvements like dark mode, charts
- OpenAI fallback for multi-model robustness
- Dockerizing the app for easy deployment

14. Conclusion

This project successfully demonstrates how AI can enrich raw lead data into personalized sales intelligence. By combining LLMs with smart UI design, I've built a scalable, practical tool for sales and marketing professionals that bridges the gap between raw contact info and meaningful outreach.

15. References

1. Groq API Documentation:
   https://console.groq.com/docs
2. Mixtral-8x7B Model Card:
   https://huggingface.co/mistralai/Mixtral-8x7B-Instruct-v0.1
3. Streamlit Documentation:
   https://docs.streamlit.io
4. Pandas Library:
   https://pandas.pydata.org/docs/
5. Python Official Documentation :
   https://docs.python.org/3/
6. OpenAI Cookbook (for prompt design references):
   https://github.com/openai/openai-cookbook