BIRD: Business Insights and Recommendations Developer using Large Language Models

Sarathbabu Karunanithi

1Affiliation not available

January 22, 2024

Abstract

Business recommendations and proposals start with asking business questions, insights from answering those questions, recommendations from those insights and final proposals or recommendations for implementation. In this paper, we present an end-to-end solution framework called BIRD (Business Insights and Recommendations Developer) which implements Large Language Models (LLMs) as a major part of this business analysis cycle in developing business questions, extracting insights, and providing recommendations in an end-to-end automated process. This framework also allows user interaction at any step for additional context or commands.
ABSTRACT

Business recommendations and proposals start with asking business questions, insights from answering those questions, recommendations from those insights and final proposals or recommendations for implementation. In this paper, we present an end-to-end solution framework called BIRD (Business Insight and Recommendation Developer) which implements Large Language Models (LLMs) as a major part of this business analysis cycle in developing business questions, extracting insights, and providing recommendations in an end-to-end automated process. This framework also allows user interaction at any step for additional context or commands.

Keywords Large Language Models · Artificial Intelligence · Business Insights · Business Recommendations

1 Introduction

In any business analysis cycle, the first step would be the stakeholder asking business related questions with the available data. The second step would be taken by data analysts or data scientists to help analyze the data to answer the questions raised by stakeholders along with generating insights. This is a continuous refinement process where there will be multiple iterations of questions asked and insights generated in a back-and-forth manner. Once all the insights are generated and business questions answered, the third step would be taken by the marketing team in coming up with recommendations or campaigns to better address these questions or to generate more revenue. In BIRD, Large Language Model (LLM) plays a key role in all these 3 steps in which it will be able to generate business questions, generate insights and generate recommendations on its own.

2 BIRD: Business Insights and Recommendation Developer

BIRD stands for Business Insights and Recommendation Developer, is a framework which takes structured data as input such as excel, csv and database files. Initial Context is obtained by parsing the columns of the data. Additional context is provided by the user about the business and the data along with explanations for each column. Using context, Business questions are generated by LLM. These business questions are generated by LLM using user provided and parsed column context. The generated business questions are answered with SQL query, python codes and general approach to the questions with the help of LLM. Generated Insights and Context are used by LLM to generate business recommendations.

BIRD framework was developed and tested with a quantized 4-bit model Llama2 [1] namely OpenOrca-Platypus2 13b from Hugging Face. BIRD was tested with an Online Retail dataset [2] in which the xlsx file was converted to SQLite which can be queried with SQL queries. BIRD framework can be applied to any structured data with any capable LLM.
Normal Business Flow

Stakeholders
Business Questions

Data Analyst
Generate Insights

Marketing Team
Recommendations

BIRD Business Flow

Data

Business Questions

Generate Insights

Recommendations

Data

Figure 1: Business Analysis Flow

Figure 2: BIRD Flow

Data: CSV, XLSX
Databases: (mysql, sql-server, etc.)

BIRD
Context Provider
Question Generator
Insight Generator
1. SQL
2. Python
3. Approach
Recommendation Generator

Figure 3: BIRD Inputs

LLM: OpenOrca Platypus2 13b (quantized 4 bit)

Input Data: Online Retail.xlsx
3 Context Provider

Large Language Model’s performance is highly influenced by the type of prompt provided to it as input. Proper prompts reduce the effect of AI Hallucination and incorrect results. Context, which is a part of prompt engineering, plays a significant role in determining the overall performance of LLMs.

Any business data that is provided to the BIRD framework is required to have proper context on what the data is about and the explanation of each column as well as the overall purpose of the data. This context provided is very important as it determines the type of business questions asked, accuracy of the insights generated, and the type of recommendation generated. BIRD requires that the columns of data be provided as context along with additional context about the data provided by the user.

4 Question Generator

After providing the user provided context along with columns of the data as prompts, LLM will be able to generate a set of business questions based on the context. If needed the user can provide additional business questions along with business questions generated by LLM.

Note: The question generated in some cases may not be relevant to the business requirements. It is up to the user to filter out irrelevant questions for better insights and recommendations.

5 Insight Generator

Insights are generated by answering the business questions asked in the previous step. User provided context along with columns of data is provided as prompts to the data along with the business question asked. Insights can be generated in one of the three ways as mentioned below,
5.1 Natural Language to SQL Query

Large Language Models will convert the business questions generated by the Question Generator along with context generated by the Context Provider to generate SQL queries related to the data. The generated SQL query is queried on the data to draw insights into each business question.

Generated Questions:
1. How many unique customers exist in the dataset?
2. What is the total sales value of all products?
3. Which customer had the highest sales value?
4. What are the top 5 selling products by quantity or revenue?
5. How does the sales distribution vary across countries?
6. Are there any particular days when a majority of transactions occur?
7. Can you predict which customers might be more likely to purchase high value items based on their past purchases?
8. Are there any discounts or promotions available in the dataset?
9. How do sales from a specific country compare to overall sales?
10. Which customer has made multiple transactions but has never spent over a certain threshold amount?

Input Data: Online Retail.xlsx
LLM: OpenOrca
Platypus2 13b (quantized 4 bit)

Generated SQL Queries:
1. SELECT COUNT(DISTINCT CustomerID) AS Unique_Customers FROM df;
2. SELECT SUM(UnitPrice * Quantity) AS TotalSalesValue FROM df;
3. .
4. .
5. .
6. .
7. .
8. .
9. .
10. SELECT CustomerID, COUNT(DISTINCT InvoiceNo) AS TransactionCount FROM df WHERE UnitPrice <= 1000 GROUP BY CustomerID HAVING TransactionCount > 1;

Generated Questions:
1. How many unique customers exist in the dataset?
2. What is the total sales value of all products?
3. Which customer had the highest sales value?
4. What are the top 5 selling products by quantity or revenue?
5. How does the sales distribution vary across countries?
6. Are there any particular days when a majority of transactions occur?
7. Can you predict which customers might be more likely to purchase high value items based on their past purchases?
8. Are there any discounts or promotions available in the dataset?
9. How do sales from a specific country compare to overall sales?
10. Which customer has made multiple transactions but has never spent over a certain threshold amount?

Figure 6: Insight Generator

Figure 7: Natural Language to SQL Query

Figure 8: Querying Database
5.2 Natural Language to Python

Large Language Models will convert the business questions generated by the Question Generator along with context generated by the Context Provider to generate Python codes related to the data [6]. The generated Python codes are executed on the data to draw insights into each business question.

![Figure 9: Natural Language to Python](image)

5.3 Approach Generator

In Case where SQL or Python are insufficient in generating answers or insights from business questions. Large Language Models will be able to convert the business questions generated by the Question Generator along with context provided by the Context Provider to generate General Approaches in solving the business question and thereby drawing insights.

![Figure 10: Approach Generator](image)

6 Recommendation Generator

Recommendation is the final step of the framework in which business recommendations are generated by using the insights drawn from previous steps by Natural Language to SQL, Python and Approach Generator along with the context provided by the Context Provider.
Note: The recommendations generated in some cases may not be relevant to the business requirements. It is up to the user to filter out irrelevant recommendations for better implementation.

Figure 11: Recommendation Generator

7 Conclusion

In this paper, we have proposed and implemented a new framework called BIRD (Business Insight and Recommendation Developer) with the help of LLMs (Large Language Models) in which we have provided the actual working of the framework in a step-by-step process along with figures. Thus, as shown with an example, BIRD can play a major role in the business analysis cycle by performing the tasks of asking business questions, answering with insights for those business questions and generating recommendations based on the insights generated in complete end to end automated process.

References