# 

# Introduction

Frequent itemset generation and rule generation are important tasks in the field of data mining, which involve finding relationships and patterns between items in a dataset. One of the most popular algorithms used for this purpose is the Apriori algorithm. Apriori is an efficient and simple algorithm for finding frequent itemsets in a dataset, which is a fundamental step in association rule mining. Association rule mining is the process of finding relationships between items in a dataset, and it is used in various fields, such as market basket analysis, web usage mining, and bioinformatics.

In this report, we will be using the Apriori algorithm to generate frequent itemsets and rules from the Dinosaur List dataset. The Dinosaur List dataset contains information about various dinosaurs, including their name, period, diet, and country of origin. This dataset was chosen because it is relevant to the topic of the report, and it can be used to demonstrate the capabilities of the Apriori algorithm in finding relationships between items.

# Method used

The following steps were taken to implement the Apriori algorithm on the Dinosaur List dataset and generate frequent itemsets and association rules.

1. **Data pre processing**: The first step was to clean and prepare the data for the analysis. The "Period" and "Diet" columns of the dataset were normalized, stripped, lowercased, and cleaned of any unwanted characters. A new dataframe was then created with data from the "Period" and "Diet" columns.
2. **Converting data**: The next step was to use the Transaction Encoder from the mlxtend library to transform the dataframe into a binary format that can be used by the Apriori algorithm. This step is necessary to convert the data into a format that the Apriori algorithm can understand.
3. **Frequent itemset generation**: The Apriori algorithm was then used to find the frequent itemsets in the encoded data. The minimum support value used was 0.1.
4. **Rule generation**: Once the frequent itemsets were generated, the association\_rules function from the mlxtend library was used to generate the association rules from the frequent itemsets. The minimum confidence value used in this experiment was 0.1.
5. **Evaluation:** The results obtained from the frequent itemset generation and rule generation were then evaluated using the support and confidence metrics. The support metric measures the number of transactions in which an itemset appears, while the confidence metric measures the degree of association between items in a rule.

# Results

The Apriori algorithm was implemented on the Dinosaur List dataset and the following results were obtained:

Graphical user interface, text, application

Description automatically generatedTable

Description automatically generated

# Conclusion

In this report, the Apriori algorithm was used to generate frequent itemsets and association rules from the Dinosaur List dataset. The results of the analysis showed that the Apriori algorithm was able to successfully find the relationships between the period and diet of the dinosaurs in the dataset.

The results showed that the Apriori algorithm is a powerful tool for finding relationships and patterns in a dataset. From the results, it can be inferred that most of the dinosaurs in the Cretaceous period were herbivorous, and the majority of the herbivorous dinosaurs were found in Jurassic period.

It should be noted that the results are limited by the sample size and quality of the data. The dataset contains only four columns and doesn't include other important information such as size, weight etc. which could be used for more complex analysis.

Overall, this report has demonstrated the use of the Apriori algorithm for frequent itemset generation and rule generation in a real-world dataset. Further research could explore the use of other algorithms, such as the FP-growth algorithm, on larger and more diverse datasets to provide more insights.

# References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. *Proceedings of the 20th international conference on Very large data bases,* 487-499.

Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *Proceedings of the 2000 ACM SIGMOD international conference on Management of data,* 1-12.

Liu, B. (2012). *Web data mining: Exploring hyperlinks, contents, and usage data.* Springer.

Rakesh, A., & Tom, M. (2014). *Apriori algorithm. International Journal of Computer Applications,* 10(9), 18-23.

Srikant, R., & Agrawal, R. (1995). Mining quantitative association rules in large relational tables. *Proceedings of the 21st international conference on Very large data bases,* 1-12.

Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data mining: practical machine learning tools and techniques* (3rd ed.). Elsevier.