**Data Analysis using Formal Concept Analysis**

*A*

***Project Report***

*submitted in partial fulfillment of the*

# **Minor Project - I**

In

Third year – Fifth Semester of

**Bachelor of Technology**

**In**

**Computer Science & Engineering**

Specialization in

# **Business Analytics and Optimization.**

*under the guidance of*

**Mr. Bikram Pratim Bhuyan, Department of Informatics**

Submitted By:

**SAP ID ROLL NO. BRANCH  
Avik Singla 500067990 R103218190 CSE-BAO**

**Sourabh Bhandari 500071057 R103218207 CSE-BAO**

**Vaishnavi Jaiswal 500071059 R103218199 CSE-BAO   
Vivek Raj 500069212 R103218189 CSE-BAO**

upes-new-logo

**Department of Informatics**

**School of Computer Science**

**University of Petroleum & Energy Studies**

**Bidholi, Via Prem Nagar, Dehradun, UK**

**November – 2020**

upes-new-logo

**DECLARATION**

I/We hereby certify that the project work entitled **“Data Analysis using Formal Concept Analysis”** in partial fulfilment of the requirements for the award of the Degree of BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING specialization Business Analytics and Optimization and submitted to the Department of Informatics, School of Computer Science, University of Petroleum & Energy Studies, Dehradun, is an authentic record of my/ our work carried out during a period from **Aug.2020** to **Dec. 2020** under the supervision of **Bikram Pratim Bhuyan**.

The matter presented in this project has not been submitted by me/ us for the award of any other degree of this or any other University.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Avik Singla** | **Sourabh Bhandari** | **Vaishnavi Jaiswal** | **Vivek Raj** |
|  | **R103218190** | **R103218207** | **R103218199** | **R103218189** |

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

**Date:** 1 December 2020  **(Name of Mentor(s))**

Bikram Pratim Bhuyan

**Dr. Thipendra Pal Singh**

Head, Department of Informatics

School of Computer Science

University of Petroleum & Energy Studies

Dehradun – 248 001 (Uttarakhand)

upes-new-logo

**ACKNOWLEDGEMENT**

We wish to express our deep gratitude to our mentor(s) **Mr. Bikram Pratim Bhuyan**, for all advice, encouragement and constant support he has given us throughout our project work. This work would not have been possible without his support and valuable suggestions.

We sincerely thank to our respected Head of Department, Informatics, **Dr. Thipendra Pal Singh,** for his guidance and support as and when required.

We are also grateful to **Dr. Manish Prateek, Dean SCS**, for providing the necessary facilities to carry out our project work successfully.

We would like to thank all our friends for their help and constructive criticism during our project work. Finally, we have no words to express our sincere gratitude to our **parents** who have shown us this world and for everything they have given to us.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Avik Singla** | **Sourabh Bhandari** | **Vaishnavi Jaiswal** | **Vivek Raj** |
|  | **R103218190** | **R103218207** | **R103218199** | **R103218189** |

**ABSTRACT**

Formal Concept Analysis is an applied branch of Lattice Theory, a mathematical discipline which enables formalization of concepts as basic units

of human thinking and analyzing data in the object-attribute form.

Now a days, data analysis has been playing a major role in various fields and required in many domains.

This work covers a detailed description of Formal Concept Analysis which covers how different processes in data analysis can be performed with the help of concepts.

The work begins by finding the concepts from the context followed by performing tasks like association rule mining, stability and clustering.

*Keywords : Data Analysis, Formal Concept Analysis, Association Rule Mining, Clustering, Stability, Concepts.*

**TABLE OF CONTENTS**

**1 Introduction 1**

**2 Related Work 2**

**3 Problem Statement 2**

**4 Objective 2**

**5 Design 3**

(Include Methodology)

(Define the Algorithm of the project)

**6 Implementation 8**

6.1 Pseudocode 8

6.2 Output Screen 15

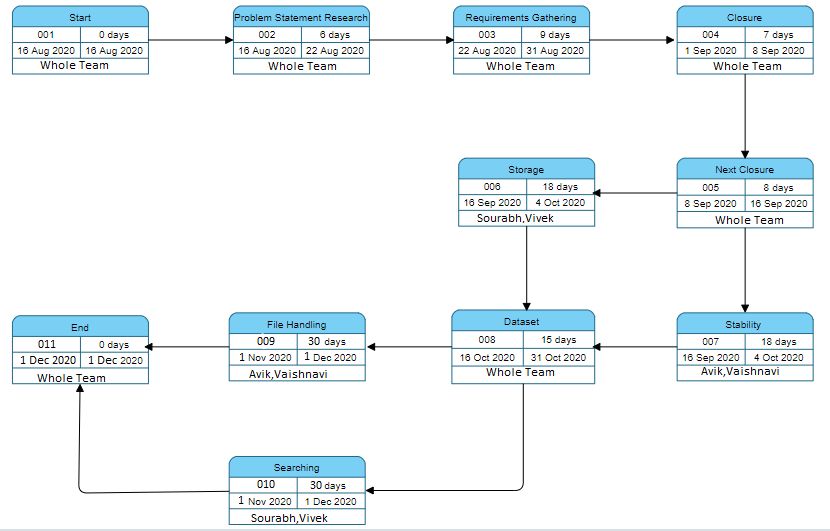
6.3 Result Analysis 8

**7 Conclusion and Future Scope 9**

**A. APPENDIX I PROJECT CODE 11**

**References**

**PERT CHART**

****

**INTRODUCTION**

When dealing with large sets of data we inevitably need to address the problem of data representation and appropriate relation among them. Many different approaches for tackling that problem have been proposed and applied over the years. One such approach is Formal Concept Analysis.

Formal concept analysis is a method of data analysis which deals with object-attribute input data. It acquires groupings, called formal concepts, from Boolean data named context, comprising of objects and attributes. The root of FCA is a formal context from which concepts are generated and are organized in a hierarchical order in the form of a lattice.

Formal Concept Analysis is concerned with the formalization of concepts and conceptual thinking and has been applied in many disciplines such as software engineering, machine learning, knowledge discovery and ontology construction during the last 20-25 years. Informally, FCA studies how objects can be hierarchically grouped together with their common attributes.

**Problem Statement**

Here we aim at analyzing a dataset from stock market Nifty 50(Top 50 companies of India listed on NSE).

We will apply the concepts and algorithms used in formal concept analysis and list out the filtered concepts according to the attributes provided.

As a result, we will get concepts that will enable us to list companies according to their previous performance and other metrics.

**Objective**

**This project will help the user to do Data Analysis on the given context using Formal Concept Analysis (FCA) and will perform following tasks: -**

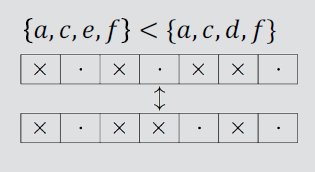
1. Study about Formal Concept Analysis: - Understanding the basic terminologies used in formal concept analysis.
2. Concept generation from given context: - Implementing algorithm to generate all concepts from the given context.
3. Prioritizing concepts: - Finding dependencies between the objects and attributes using stability index.
4. Clustering :- Arranging objects into groups according to their similar attributes.
5. Storing and Searching of Concepts using Linked Lists.

**Design**

**Methodology:**

We will begin by taking stock market dataset and ruling out concepts which will in turn help the users to find a suitable company to invest in according to their needs.

**ALGORITHMS:**

****ALGORITHM 1: All closures ( M , ’’ ) : generating all closed sets  
Input: A closure operator X->X’’ on a finite linearly ordered set M.  
Output: All closed sets in lectic order.  
A = FirstClosure( ” )  
while A =! NULL do  
Output A  
A := NextClosure ( A , M , ’’ )

Lectic Order

ALGORITHM 2: First Closure ( ’’ )  
Input: A closure operator X->X” on a finite linearly ordered set.  
Output: The lectically first closed set, i.e., the closure of the empty set.  
return 0

ALGORITHM 3: Next Closure ( A , M , ” )   
Input: A closure operator X->X” on a finite linearly ordered set M and a subset A ∈ M.  
Output: The lectically next closed set after A if it exists; NULL otherwise.  
For all m ∈ M in reverse order do  
if M ∈ A then  
A := A – {m}  
else  
B := (A U {m})’’  
If B-A contains no element < m then  
Return B  
return NULL

Algorithm 4: Stability

Finding Stability index for filtering.

Input :- Set of attribute A, set of object B and Context

Output :- Stability index for given concept

num=0,dem=2^M

For all C, where C is all possible subsets of M:

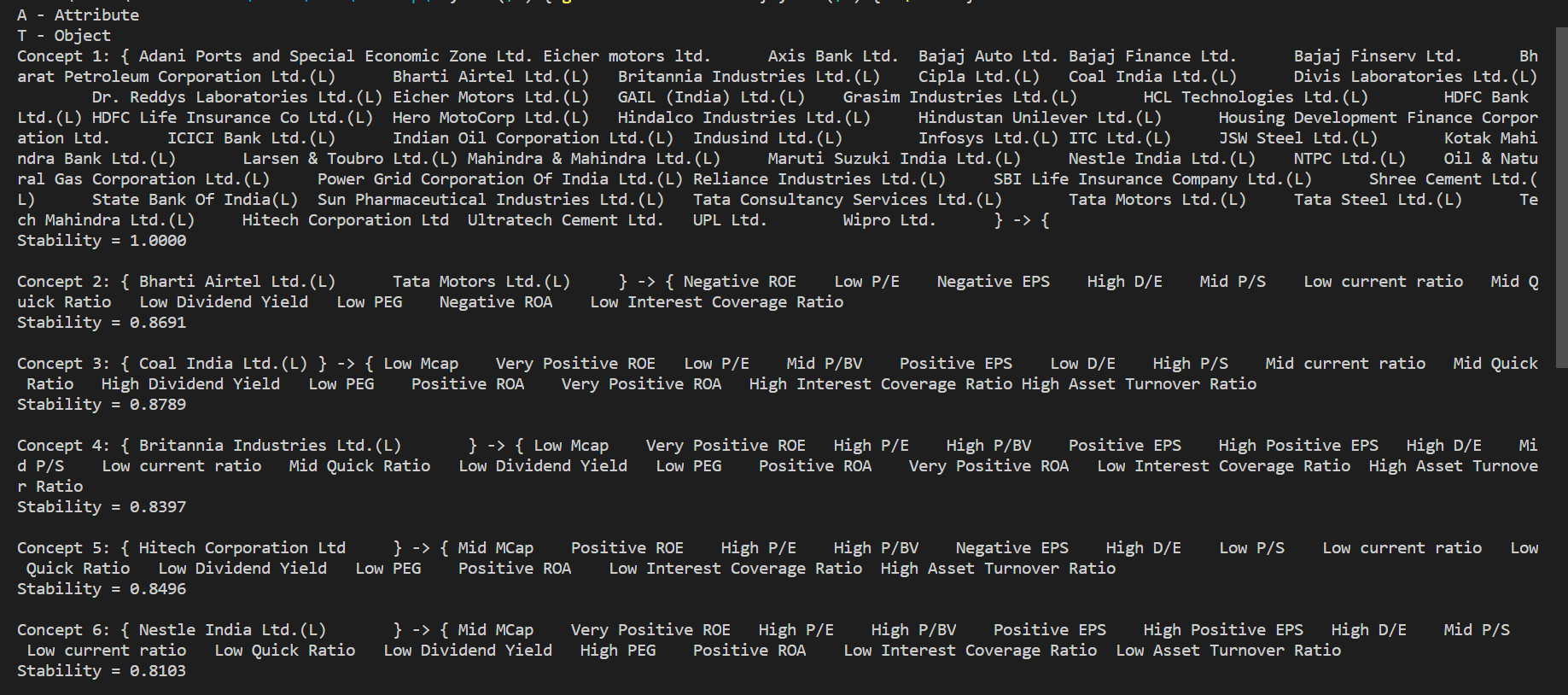
if derivative of C =B

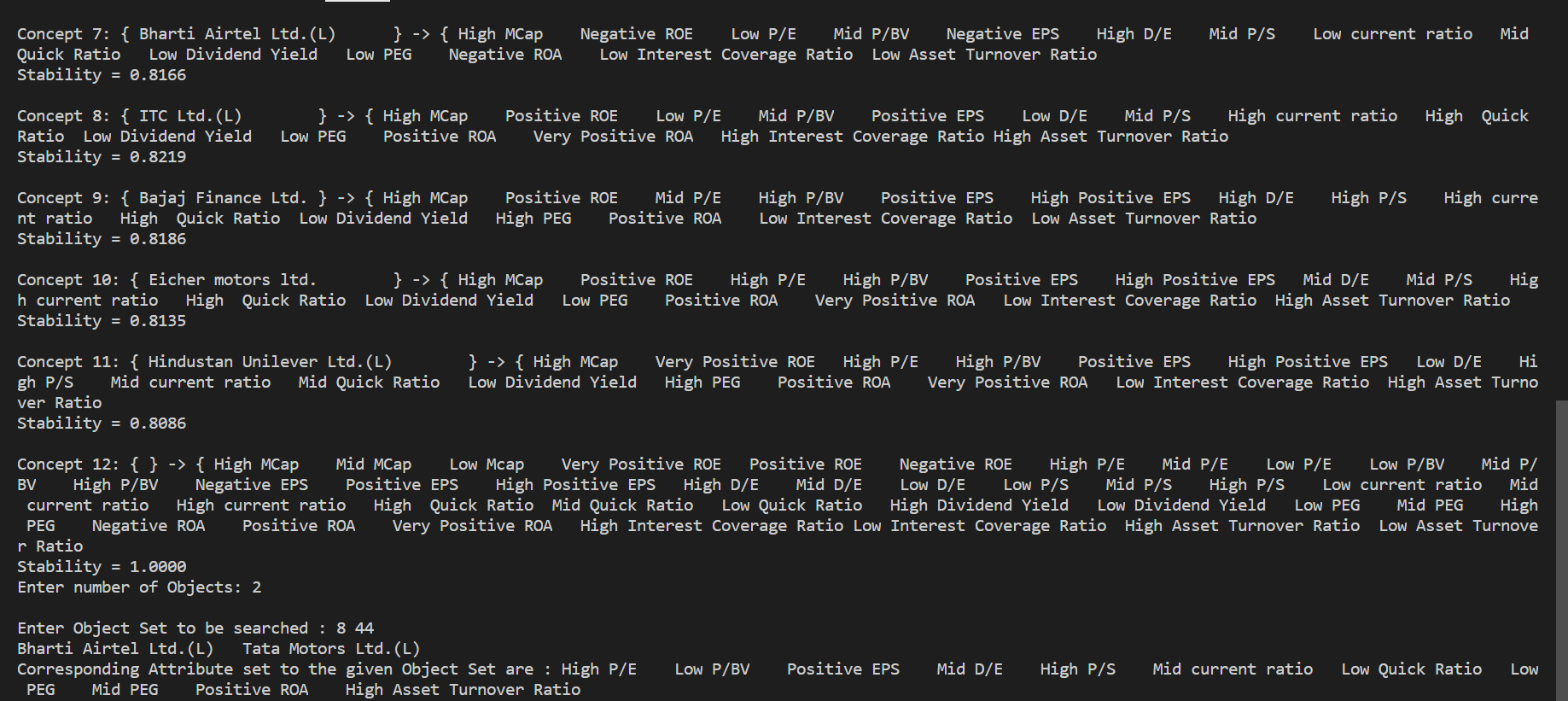
num++

return num/dem

**Implementation**

Output Screen:





**Result Analysis**

Here, in the above output we can see that we are generating concepts based on object-attribute relations. We used stability and threshold value to filter out concepts or the companies best fit according to different financial ratios (attributes) to get a list of suitable companies for our investors.

**Conclusion and Future Scope**

By doing this project on Formal Concept Analysis we learnt about exploring various object-attribute relationships in a context. We found out useful relations between objects and their attributes to derive meaningful conclusions and stability of the relations.

Further, in future we shall extend our work to build a stock market portfolio screener which aims at screening the stocks listed on stock market exchange and telling the end users about certain companies best fit for their investment needs and goals.

**APPENDIX I PROJECT CODE**

The entire project code and its implementation is managed on GitHub by us at:

[**vivekrajx/Data-Analysis-using-Formal-Concept-Analysis: This repository includes our work for Minor Project of Sem-5. (github.com)**](https://github.com/vivekrajx/Data-Analysis-using-Formal-Concept-Analysis)

**References**

[1] <https://www.coursera.org/learn/formal-concept-analysis> , Sergei Obiedkov,

National Research University Higher School of Economics.

[2] <https://www.upriss.org.uk/> Uta Priss, 2006.

Algorithms used:

Next Closure, Stability from

[3] Dmitry I. Ignatov: Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Fields, National Research University Higher School of Economics, Moscow, 2015. <https://bit.ly/2JpnYWf>

[4] <https://en.wikipedia.org/wiki/Formal_concept_analysis>

[5] Financial Information for building dataset:

* https://moneycontrol.com
* <https://finviz.com>
* <https://screener.in>