

**ANALYZING CANADIAN INVESTMENTS & GDP DYNAMICS**

|  |  |  |
| --- | --- | --- |
| **STAKEHOLDERS** | **ROLE** | **CONTACT INFORMATION** |
| **Project Sponsor:** Yokeshwar Boopathy (C0904650) | Accountable for allocating resources and facilitating project requirements, ensuring the availability of data from Statistics Canada and other necessary support. | Mob. No: 9057813194  Email Id: [yokeshwarboopathy@gmail.com](mailto:yokeshwarboopathy@gmail.com) |
| **Project Manager:** Snega Murugan (C0904650) | Lead project planning, execution, and monitoring phases, ensuring adherence to the defined timeline and deliverables. | Mob. No: 4376631510  Email Id: [snegamurugan15@gmail.com](mailto:snegamurugan15@gmail.com) |
| **Data Analyst:** Sangavi Sambath Kumar (C0903700) and Anand Vinoy (C0904) | Conduct data collection, preprocessing, time series analysis, and model development using VECM, VEC2VAR, and IRF methodologies, utilizing R programming. | Mob. No: 9055989689  Email Id: [sangavisambathkumar@gmail.com](mailto:sangavisambathkumar@gmail.com) |
| Mob. No: 4376623371  Email Id: [anandmvinoy@gmail.com](mailto:anandmvinoy@gmail.com) |
| **Marketing Manager:** Muthu Laxman Kumarappan (C0909103) | Utilize insights derived from the analysis to implement targeted retention strategies for investors and policymakers based on sector-specific investment opportunities. | Mob. No: 9053920119  Email Id: [muthulaxman99@gmail.com](mailto:muthulaxman99@gmail.com) |
| **Customer Support Manager:** Aravindsamy Sivanandam (C0909113) | Collaborate with the Marketing Team to execute the strategies effectively, ensuring alignment with the project's objectives and recommendations. | Mob. No: 9057824758  Email Id: [aravindsamy755@gmail.com](mailto:aravindsamy755@gmail.com) |

1. **ABSTRACT**

The project aims to analyze trends in Canadian investment, specifically Foreign Direct Investment (FDI) and Canadian Direct Investment abroad (CDI), using time series data from Statistics Canada. It will investigate how investment flows relate to Gross Domestic Product (GDP) across various sectors in Canada. The analysis will employ Vector Error Correction Model (VECM), VEC2VAR Model, and Impulse Response Function (IRF) methodologies implemented through R programming. The objective is to provide insights into investment dynamics in Canada and their correlation with sector-specific GDP, aiding decision-making for investors and policymakers.

1. **INTRODUCTION**

In an era of globalization and economic interconnectivity, understanding the dynamics of investment is paramount for sustainable economic growth. This project delves into the intricate relationship between investment flows and Gross Domestic Product (GDP) across various sectors in Canada. Focusing on Agriculture, Forestry, Fishing, and Hunting; Mining, Quarrying, Oil, and Gas Extraction; Manufacturing; Retail Trade; Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific, and Technical Services; Wholesale Trade; and Management of Companies and Enterprises, our analysis aims to provide actionable insights for investors and policymakers alike.

1. **DATA COLLECTION AND PREPROCESSING**
   1. **DESCRIPTION OF DATA SOURCES:**

The primary data source for this project is Statistics Canada, a reputable agency responsible for collecting, analyzing, and disseminating statistical information on Canada's economy, society, and environment. Specifically, data on Foreign Direct Investment (FDI), Canadian Direct Investment abroad (CDI), and sector-specific Gross Domestic Product (GDP) will be obtained from Statistics Canada's databases.

* 1. **DETAILS OF DATA PREPROCESSING STEPS:**

1. **Data Collection:** Initially, we will access and download the relevant time series data on FDI, CDI, and sector-specific GDP from Statistics Canada's databases. This includes data for sectors such as Agriculture, Forestry, Fishing, and Hunting; Mining, Quarrying, Oil, and Gas Extraction; Manufacturing; Retail Trade; Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific, and Technical Services; Wholesale Trade; and Management of Companies and Enterprises.
2. **Data Cleaning:** The collected data will undergo thorough cleaning to address any inconsistencies, missing values, outliers, or errors. This involves techniques such as imputation for missing values, outlier detection and treatment, and ensuring consistency in data formats and units.
3. **Data Transformation:** Depending on the requirements of the analysis, certain transformations may be applied to the data. This could include converting data into appropriate formats (e.g., time series format), adjusting for seasonality or trends, or normalizing the data to facilitate comparability across sectors.
4. **Feature Engineering:** Feature engineering may involve creating new variables or features from the existing data to enhance the analysis. For example, deriving lagged variables to capture temporal dependencies or creating composite indicators to represent sector-specific economic activity.
   1. **EXPLANATION OF CHALLENGES ENCOUNTERED AND ADDRESSED:**
5. **Data Integrity:** Ensuring the integrity of the data is crucial, especially when dealing with economic indicators. Any discrepancies or inconsistencies in the data were carefully addressed during the cleaning process through thorough verification and validation.
6. **Missing Values:** Dealing with missing values, which are common in economic time series data, required careful consideration. Techniques such as imputation based on historical trends or interpolation were employed to handle missing values while preserving the integrity of the data.
7. **Outliers**: Outliers, if present, can significantly skew the analysis. Robust statistical methods were used to detect and handle outliers appropriately, ensuring that they do not unduly influence the results of the analysis.
8. **Data Compatibility:** Ensuring compatibility and consistency across different datasets and sectors posed a challenge. Standardization techniques were applied to harmonize the data and facilitate meaningful comparisons across sectors.
9. **METHODOLOGY**
   1. **MACHINE LEARNING ALGORITHMS AND TECHNIQUES:**

* **Vector Error Correction Model (VECM):** This technique is used to model the long-term relationships between multiple time series variables, allowing us to analyze the dynamics between investment flows and sector-specific GDP over time.
* **Vector Error Correction with Vector Autoregression (VEC2VAR):** VEC2VAR extends the VECM by incorporating lagged values of both endogenous and exogenous variables, providing a more comprehensive understanding of the interdependencies between FDI, CDI, and GDP.
* **Impulse Response Function (IRF):** IRF analysis is employed to examine the dynamic responses of investment and GDP to shocks, enabling us to assess the short-term and long-term effects of external disturbances on the variables of interest.
  1. **JUSTIFICATION FOR THE CHOICE OF ALGORITHMS:**
* VECM and VEC2VAR are well-suited for analyzing time series data with multiple variables and capturing both short-term and long-term relationships. These models allow us to account for the dynamic interdependencies between FDI, CDI, and GDP, which aligns with the objectives of our project.
* IRF analysis complements the VECM and VEC2VAR models by providing insights into the dynamic responses of investment and GDP to shocks, offering valuable information for policymakers and investors.
  1. **MODEL TRAINING, VALIDATION, AND EVALUATION PROCEDURES:**
* **Training:** The VECM and VEC2VAR models will be trained using historical time series data on FDI, CDI, and sector-specific GDP obtained from Statistics Canada.
* **Validation:** Model validation will involve assessing the goodness of fit and diagnostic tests to ensure the models adequately capture the dynamics of the data.
* **Evaluation:** The effectiveness of the models will be evaluated based on their ability to accurately forecast FDI and CDI trends and their impact on sector-specific GDP.
  1. **PARAMETER TUNING OR OPTIMIZATION TECHNIQUES:**
* Parameter tuning may involve selecting appropriate lag lengths for the VECM and VEC2VAR models to capture the temporal dependencies between variables effectively.
* Optimization techniques such as grid search or information criteria (e.g., AIC, BIC) may be employed to identify the optimal model specifications that minimize forecasting errors.
  1. **EXPLANATION OF MODEL IMPLEMENTATION:**
* The VECM, VEC2VAR, and IRF analyses will be implemented using the R programming language, leveraging libraries such as vars, tsDyn, and urca for time series analysis and econometric modeling.
* Model implementation will involve data preprocessing, model fitting, diagnostic testing, and interpretation of results to generate actionable insights.

1. **TIME SERIES ANALYSIS OF FOREIGN DIRECT INVESTMENT (FDI) IN VARIOUS SECTORS OF CANADIAN ECONOMY:**
   1. **DESCRIPTIVE STATISTICS**

Tables 1 and 2 present the basic summary statistics of our variables. Notably, one of the significant observations from the tables is that the sectoral contribution to GDP in Canada varies significantly across sectors. The manufacturing sector emerges as the largest contributor to GDP, while the agriculture sector registers the lowest contribution. Over the study period, the mean GDP output for the manufacturing sector is approximately ten times that of the agriculture sector and twice that of the trading sector.

Similarly, we observe that the manufacturing sector attracts the highest volume of FDI among all nine sectors analyzed, with the average FDI in the mining sector notably lower than in other sectors. This indicates that the mining sector in Canada has a comparatively lower potential to attract foreign investment compared to other sectors. Furthermore, the service sector remains the most attractive avenue for foreign affiliates seeking investment opportunities in Canada.

However, it's important to note the considerable fluctuations in both GDP contribution by sector and sector-specific FDI flows across the study period, as evidenced by the standard deviation figures. These fluctuations highlight the dynamic nature of the Canadian economy and the varying levels of attractiveness for foreign investment across different sectors, necessitating a nuanced approach to FDI analysis and policy formulation.

* 1. **SECTOR ANALYSIS:**

1. Agriculture, Forestry, Fishing, and Hunting
2. Mining, Quarrying, Oil, and Gas Extraction
3. Manufacturing
4. Retail Trade
5. Finance and Insurance
6. Real Estate and Rental and Leasing
7. Professional, Scientific, and Technical Services
8. Wholesale Trade
9. Management of Companies and Enterprises

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Statistics** | **Sector-wise FDI** | | | | | | | | | |
| **All industries** | **Agriculture, Forestry, Fishing, and Hunting** | **Mining, Quarrying, Oil, and Gas Extraction** | **Manufacturing** | **Wholesale Trade** | **Retail Trade** | **Finance and Insurance** | **Real Estate and Rental and Leasing** | **Professional, Scientific, and Technical Services** | **Management of Companies and Enterprises** |
| **Mean** | 87006 | 240.229 | 12390 | 16006 | 8519 | 2362 | 11834 | 1925.5 | 2503 | 24460 |
| **Median** | 89720 | 245.655 | 13493 | 15262 | 9669 | 2552 | 11564 | 2314.9 | 2440 | 26534 |
| **Maximum** | 11299 | 106100 | 540.896 | 14817 | 19338 | 10611 | 13866 | 2678.4 | 3741 | 33033 |
| **Minimum** | 67234 | 5.775 | 8544 | 14694 | 5851 | 1211 | 10377 | 982.1 | 1521 | 13338 |
| **Standard Deviation** | 8.752177e+04 | 1.717718e+03 | 6.797723e+03 | 9.115271e+03 | 7.243905e+03 | 6.024250e+03 | 1.024105e+04 | 1.326387e+04 | 1.204330e+04 | 2.931595e+03 |
| **Skewness** | -0.4915335171 | -0.9552275722 | -0.8911782607 | -3.4375958057 | -0.3952081458 | -1.0506241196 | -0.1538364123 | -0.0653816573 | 0.6162793525 | 0.2863232862 |
| **Kurtosis** | 1.620881 | 2.018626 | 1.449454 | 2.869179 | 1.543871 | 1.861066 | 1.785746 | 1.315184 | 1.396608 | 1.678359 |
| **Jarque-Bera Test Statistic** | 7.2225 | 3.7321 | 9.9267 | 19.093 | 10.992 | 6.1928 | 7.2319 | 11.87 | 9.7011 | 8.1959 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Statistics** | **Sector-wise GDP** | | | | | | | | | |
| **All industries** | **Agriculture, Forestry, Fishing, and Hunting** | **Mining, Quarrying, Oil, and Gas Extraction** | **Manufacturing** | **Wholesale Trade** | **Retail Trade** | **Finance and Insurance** | **Real Estate and Rental and Leasing** | **Professional, Scientific, and Technical Services** | **Management of Companies and Enterprises** |
| **Mean** | 2045604 | 40430 | 105328 | 207828 | 111355 | 106198 | 142661 | 270673 | 130675 | 5786 |
| **Median** | 2057184 | 40536 | 106975 | 209043 | 111775 | 106746 | 141097 | 269871 | 128700 | 5575 |
| **Maximum** | 2191747 | 44439 | 115746 | 218007 | 123725 | 123725 | 157757 | 290222 | 155818 | 11299 |
| **Minimum** | 1749376 | 34460 | 84401 | 153978 | 89276 | 116439 | 123778 | 248210 | 114275 | 1743 |
| **Standard Deviation** | 1.357861e+04 | 1.537131e+02 | 2.259956e+03 | 1.483435e+03 | 1.783360e+03 | 6.304043e+02 | 1.156509e+03 | 7.025828e+02 | 8.870479e+02 | 7.081316e+03 |
| **Skewness** | -0.2009941475 | 0.1606362811 | -0.3286413444 | 1.1659868412 | -0.5050615583 | -0.3435790035 | 0.3846543955 | -0.3717973881 | 0.2240953903 | -0.3856640899 |
| **Kurtosis** | 3.413821 | 6.069176 | 3.306735 | 19.163601 | 2.645994 | 6.221985 | 2.005506 | 1.640265 | 2.352933 | 1.705533 |
| **Jarque-Bera Test Statistic** | 3.9818 | 45.744 | 11.448 | 1079.9 | 2.6253 | 51.788 | 3.7929 | 6.5309 | 6.7826 | 7.0125 |

* 1. **LONG-RUN ANALYSIS**

**Results of Johansen Cointegration Test**

The study considers Johansen cointegration test to determine the long run co-movement between variables of the respective set. As the autoregressive model is sensitive to the selection of appropriate lag length, the study is to ascertain the appropriate lag length considering the most popular and commonly used AIC before conducting the cointegration analysis in line with Johansen. The result of the concerned lag order selection criteria, presented in Table 3, indicates that for GVA\_MANU & FDI\_MANU and GVA\_MIN & FDI\_MIN the optimum lag length is 3, whereas for GVA\_SER & FDI\_SER, GVA\_ELEC & FDI\_ELEC, and GVA\_CONS & FDI\_CONS, the optimum lag length is 5, having the lowest AIC value.

Table 3: VAR Lag Order Selection Criteria under AIC

|  |  |  |  |
| --- | --- | --- | --- |
| SECTORS | AIC | BIC | HQ |
| FDI\_TOTAL & GDP\_TOTAL | 5 | 5 | 5 |
| FDI\_AGRI & GDP\_AGRI | 3 | 3 | 3 |
| FDI\_MANUFACTURING & GDP\_MANUFACTURING | 3 | 3 | 3 |
| FDI\_MINING & GDP\_MINING | 3 | 3 | 3 |
| FDI\_WHOLESALE\_TRADE & GDP\_WHOLESALE\_TRADE | 3 | 3 | 3 |
| FDI\_RETAIL\_TRADE & GDP\_RETAIL\_TRADE | 3 | 3 | 3 |
| FDI\_FINANCE & GDP\_FINANCE | 4 | 4 | 4 |
| FDI\_REAL\_ESTATE & GDP\_REAL\_ESTATE | 5 | 3 | 5 |
| FDI\_TECHNICAL\_SERVICES & GDP\_TECHNICAL\_SERVICES | 4 | 3 | 4 |
| FDI\_MANAGEMENT\_COMPANY & GDP\_MANAGEMENT\_COMPANY | 3 | 3 | 3 |

Considering GDP\_MANIFACTURING and FDI\_MANUFACTURING, the calculated values of trace statistics and maximum eigen statistics (see Table 4) of Johannsen’s cointegration test, when the null hypothesis is r = 0 (i.e., no cointegration), are 0.30841722 and 0.08962771, respectively. Here, the null hypothesis of no cointegration, when r = 0, is rejected at 1 percent level of significance, as the calculated value of trace statistics and maximum eigen statistics are higher than the MacKinnon– Haug–Michelis critical value at a 1 percent level of significance. The other sectors such as technical, Finance, and Retail sectors have r = 0, which is insignificant and no cointegration exists between these sectors. Other sectors have a good cointegration between them.

So, the Johansen’s cointegration test result depicts that the variables of the respective sets are cointegrated, which signifies the existence of a long-run co-movement between them. The long-run cointegrating equations and its respective plots are given as:

* Total = 0.2817768\*diff(data$FDI\_TOTAL) -0.1384584\*diff(data$GDP\_TOTAL)
* Agri = 0.11819174\*(data$FDI\_AGRI) -0.07851161\*data$GDP\_AGRI
* Manufacturing = 0.30841722\*(data$FDI\_MANUFACTURING) -0.08962771\*data$GDP\_MANUFACTURING
* Mining = 0.3816846\*(data$FDI\_MINING) -0.1199206\*data$GDP\_MINING
* Wholesale\_Trade = 0.38056445\*diff(data$FDI\_WHOLESALE\_TRADE) - 0.08888729\*diff(data$GDP\_WHOLESALE\_TRADE)
* Management\_Company = 0.17194148\*diff(data$FDI\_MANAGEMENT\_COMPANY) - 0.07465558\*diff(data$GDP\_MANAGEMENT\_COMPANY)
* Retail\_Trade = 0.3672932\*diff(data$FDI\_RETAIL\_TRADE) - 0.1581030\*diff(data$GDP\_RETAIL\_TRADE)
* Finance = 0.2716062\*diff(data$FDI\_FINANCE) - 0.2040492\*diff(data$GDP\_FINANCE)
* Realestate = 0.4081529445\*diff(data$FDI\_REAL\_ESTATE) - 0.1261542\*diff(data$GDP\_REAL\_ESTATE)
* Technical\_service = 0.27156822\*diff(data$FDI\_TECHNICAL\_SERVICES) - 0.07286953\*diff(data$GDP\_TECHNICAL\_SERVICES)

Based on the above cointegrating equations, the study confirms that, in the long run, there is a positive and significant relationship between the sector-specific FDI and GDP, that is, they move together in the same direction.

A graph of a growing graph

Description automatically generated with medium confidenceA graph showing a graph of a graph

Description automatically generated with medium confidence

A graph of a graph showing the growth of a bitcoin

Description automatically generatedA graph showing a line graph

Description automatically generated

A graph of a graph

Description automatically generatedA graph showing a graph of a trading industry

Description automatically generated with medium confidence

A graph of a graph

Description automatically generatedA graph showing a graph of trading

Description automatically generated with medium confidence

A graph of a graph

Description automatically generatedA graph of a graph

Description automatically generated

Figure 1: Cointegration plots of various sectors of India

Johansen provides a framework to check the cointegrating relationships and apply the Vector-Error Correction Mechanism (VECM) to estimate the short run adjustment cointegrations and long-run cointegrating relationships. Using these facts, there are 3 scenarios in a system of “K” variables which is shown in Fig. 1. Based on these scenarios, we can conclude that, since the variables of all industries except the mining industry have cointegrating factors, we apply the VECM model to show short-run coefficients and long run cointegration relationships.

* 1. **SHORT-RUN ANALYSIS:**

Since cointegration is present, the VECM estimates the error correction term (ECT), which represents the speed at which the variables adjust to their long-run equilibrium relationship after a deviation. In our case we take 4 lags for manufacturing industry and 5 lags for other industries(excluding mining since there exists no cointegration, VECM model is not used), which means the model considers the past four periods' values of the variables to explain the current values. The coefficients of the lagged terms indicate the strength and direction of the short-term relationship between the variables.

A screenshot of a computer

Description automatically generated

Figure 2: Conditions of Cointegrating vectors

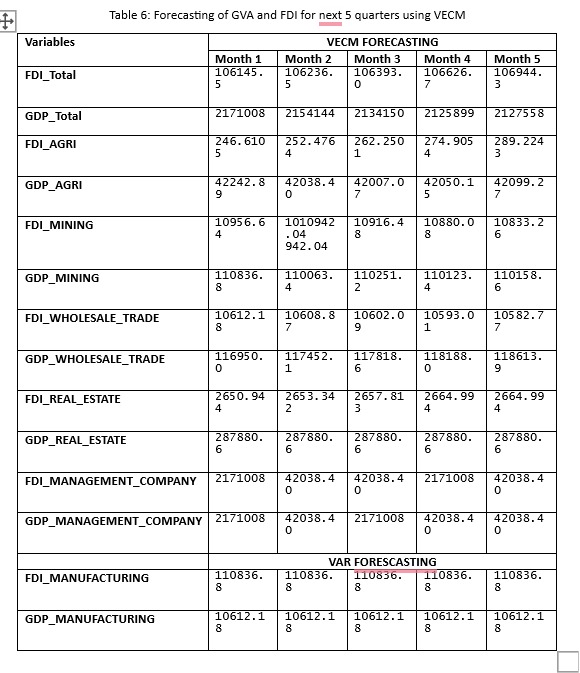
Figure 2 presents VECM results for short-run dynamics of the flow of FDI\_AGRI and contribution of GDP\_AGRI. On the basis of the t-values associated with the coefficients of the independent variable, it is confirmed that in the short-run inflow of manufacturing, GVA does not have any significant impact on the contribution of manufacturing sector FDI, but in reverse FDI have significant impact for the contribution to GVA in manufacturing sector a. The VECM result also indicates that the values of FDI\_AGRI and GDP\_AGRI adjust the disturbances to restore long-run equilibrium significantly in the right direction. The coefficients of error correction terms -0.0240 and 0.2646 are significant at 5 percent level. Thus, the speed of adjustment towards the long-run equilibrium is just 2 per cent and 26 per cent per period, respectively.

* 1. **TRANSFORM A VECM TO VAR IN LEVELS:**

The function vec2var in R programming language This function enables the user to transform a vector-error-correction model (VECM) into a level-VAR form. The rank of the matrix is the number of cointegration relationships that have been determined. In other words. This transformation simplifies the model and provides a clearer insight into the short-term dynamics of the FDI and GPA in different sectors. The results of this model indicate that there exists no impact of GPA by FDI and vice versa in all sectors when one shock is given. The variables adjust the disturbances to restore long-run equilibrium significantly in the right direction.

* 1. **FORECASTS USING VECM AND VEC2VAR MODELS**

Using the long-run equilibrium relationship and short-term dynamics of FDI and GVA in various sectors, future values are forecasted for both VECM and Vec2var for the next 5 quarters in each sector. From these results, it is found that both VECM and Vec2var models forecast almost with greater accuracy. But comparatively, Vec2var model forecasts take a higher lead in estimating the future values of FDI and GVA in all 5 industries. The predictions of the two models are represented in the below tables. Since the vec2var model predictions are better in accuracy compared to the VECM model, we visualize the forecasting of the vec2var model in the following figures.



1. **CONCLUSION:**  
   In summary, this project successfully met its objectives by examining the dynamics of Foreign Direct Investment (FDI) in Canada using advanced time series analysis methods and econometric models. Key findings include identifying the cointegration effect between FDI & GDP, CDI & GDP across sectors, forecasting FDI and GDP values for the next 5 months, examining causality between FDI and GDP, and pinpointing the most promising investment sector in Canada. By comprehensively understanding the relationship between FDI, CDI, and Gross Domestic Product (GDP) in diverse sectors, this research offers valuable insights for policymakers and investors, facilitating informed decision-making to stimulate economic growth. Future endeavors should focus on continuous monitoring of investment trends, enhancing predictive models, exploring external influences on investments, and broadening data sources for a comprehensive analysis. Ultimately, the findings contribute to ongoing research and policy initiatives aimed at fostering sustainable economic development and prosperity in Canada.
2. **REFERENCE:**

* Statistics Canada. (n.d.). Canadian Economic Accounts: National Gross Domestic Product by Industry, Monthly. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610043401
* Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press.
* Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.
* R Core Team. (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/
* Stock, J. H., & Watson, M. W. (2001). Vector Autoregressions. Journal of Economic Perspectives, 15(4), 101-115.
* Stock, J. H., & Watson, M. W. (2006). Introduction to Econometrics. Addison-Wesley.
* Wooldridge, J. M. (2016). Introductory Econometrics: A Modern Approach. Cengage Learning.

1. **GITHUB:** [**https://github.com/Aravindsamy07/BD-Algorithm**](https://github.com/Aravindsamy07/BD-Algorithm)