

Low Level Design (LLD)

InvestmentAnalysis

Revision Number - 1.2
Last Date of Revision–16/02/2023

SIDDHARTH JAIN

Document Control

Date	Version	Description	Author
10/02/2023	1.0	Introduction, Problem Statement	Siddharth Jain
12/02/2023	1.1	Dataset Information, Architecture Description	Siddharth Jain
16/02/2023	1.2	Final Revision	Siddharth Jain

Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Investment Analytics prediction analysis technique. It will explain the necessary steps which have to be followed before any analysis can begin. LLD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The LLD will be focusing on the below objectives:

- Problem Understanding.
- DataAcquisition.
- Data Pre-Processing and ExploratoryAnalysis
- Development ofmodels
- Auditing accuracy and retrain ifrequire
- Finalizing themodel
- Dashboard report for importantactivities

Scope

The LLD documentation presents the detailed structure of the Investment analytics for each of its individual components. The goal of LLD is to give the internal logical design of the actual program code. Low-level design is created based on the high-level design. The LLD documentation contains the complete description of the model used along with the comparisons of the proposed model/library compared with a baseline(existing) model against a set of metrics.

Project Introduction

The practice of evaluating an investment for profitability and risk is known as investment analysis. Its ultimate goal is to determine whether a certain investment is a good fit for a portfolio. It can also range from a single bond in a personal portfolio to a fledgling business investment and even large-scale corporate ventures.

Investment is a game of understanding historic data of investment objects under different events but it is still a game of chances to minimize the risk we apply analytics to find the equilibrium investment.

The given dataset contains sector and financial year-wise data of FDI in India.

Constraints

Our analysis is done based on a limited dataset provided for different 63 sectors and 17 years investment. The analysis is done sector and year wise.

Risks

Document specific risks that have been identified or that should be considered.

Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

1. Technical specifications

Dataset

The Dataset is taken from iNeuron's provided dataset-

```
#Importing the given dataset
df = pd.read_csv(r"C:\Users\hp\Desktop\inueron internship\3)InvestmentAnalysis\1.My project\3.Code\Dataset.csv")
df
```

	Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17
0	METALLURGICAL INDUSTRIES	22.69	14.14	36.61	8.11	200.38	149.13	169.94	1175.75	959.94	419.88	1098.14	1786.14	1466.23	567.63	359.34	456.31	1440.18
1	MINING	1.32	6.52	10.06	23.48	9.92	7.40	6.62	444.36	34.16	174.40	79.51	142.65	57.89	12.73	684.39	520.67	55.75
2	POWER	89.42	757.44	59.11	27.09	43.37	72.69	157.15	988.68	907.66	1271.79	1271.77	1652.38	535.68	1066.08	707.04	868.80	1112.98
3	NON-CONVENTIONAL ENERGY	0.00	0.00	1.70	4.14	1.27	1.35	2.44	58.82	125.88	622.52	214.40	452.17	1106.52	414.25	615.95	776.51	783.57
4	COAL PRODUCTION	0.00	0.00	0.00	0.04	0.00	9.14	1.30	14.08	0.22	0.00	0.00	0.00	0.00	2.96	0.00	0.00	0.00
...
58	PRINTING OF BOOKS (INCLUDING LITHO PRINTING IN...	0.00	0.00	6.30	0.00	0.06	9.90	20.04	35.54	31.61	70.51	36.63	47.39	14.34	113.78	72.58	122.81	53.17
59	COIR	0.00	0.00	0.00	0.00	0.47	0.59	0.04	0.01	0.00	0.25	0.10	0.55	0.15	0.54	1.36	0.00	0.00
60	CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES	0.00	0.00	0.00	0.00	0.00	0.93	64.06	182.92	172.70	324.56	675.07	386.28	283.89	485.37	870.25	4510.71	1860.73
61	CONSTRUCTION DEVELOPMENT: Townships, housing, ...	24.33	51.75	36.10	47.04	152.06	228.71	1392.95	3887.33	4657.51	5466.13	1663.03	3140.78	1332.49	1226.05	769.14	112.55	105.14
62	MISCELLANEOUS INDUSTRIES	832.07	221.37	218.76	235.48	121.83	164.76	304.87	528.42	1549.70	1147.56	1475.97	813.38	229.49	468.74	765.88	668.77	296.40

63 rows x 18 columns

Figure 1: Investment Analytics Dataset

The dataset consists of 63 individual data. There are 18 columns in the dataset which are described below.

1. Sector – There are total 63 sector's name. Some of them are
 - a. METALLURGICAL INDUSTRIES
 - b. MINING
 - c. POWER
 - d. NON-CONVENTIONAL ENERGY
 - e. COAL PRODUCTION etc.
2. In the 1st Column Sector name is mentioned and then rest in 17 columns from 2000-01 to 2016-17 historic data of investment are mentioned for the mentioned sectors.

3. Problem Statement

Investment is a game of understanding historic data of investment objects under different events but it is still a game of chances to minimize the risk we apply analytics to find the equilibrium investment.

To understand the Foreign direct investment in India for the last 17 years from 2000-01 to 2016-17. This dataset contains sector and financial year-wise data of FDI in India

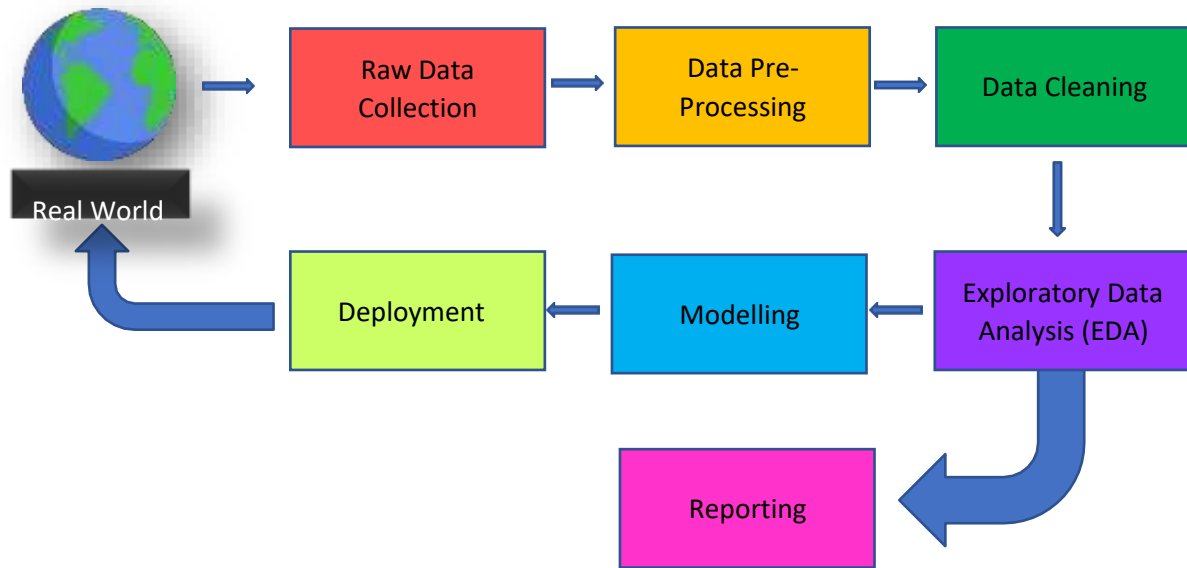
Sector-wise investment analysis

Year-wise investment analysis

Find key metrics and factors and show the meaningful relationships between attributes.

Do your own research and come up with your findings

4. Architecture



4.1 ArchitectureDescription

- 1. Raw Data Collection-** The Dataset was taken from iNeuron provided Project Description Document.

<https://drive.google.com/drive/folders/1M5z7z1NmWar7y1eFs67orfjqHL0iSViL?usp=sharing>

- 2. Data Pre-Processing**

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data to the model to train.

This Process includes-

- a) Handling Null/Missing Values

- 3. Data Cleaning**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

- a) Remove duplicate or irrelevant observations
- b) Filter unwanted outliers
- c) Renaming required attributes

- 4. Exploratory Data Analysis (EDA)**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis

and to check assumptions with the help of summary statistics and graphical representations.

5. Reporting

Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in easy in report because your model will be used by many stakeholders who are not from technical background.

- High Level Design Document(HLD)
- Low Level Design Document(LLD)
- Architecture
- Wireframe
- Detailed ProjectReport
- Power PointPresentation

6. Modelling

Data Modelling is the process of analysing the data objects and their relationship to the other objects. It is used to analyse the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

7. Deployment

Tableau dashboard

