



PG Diploma in ML

Lecture On : Investment
Case Study

Instructor : Arihant Jain

Today's Agenda

- 1 Discussing the Premise of the Assignment
- 2 The objectives for the assignment
- 3 Pre-Modelling Steps
- 4 Analysis Steps and Checkpoints
- 5 Q & A

- Data exploration
 1. Loading data using encoding ISO-8859-1.
Try using different encoding formats.
chardet library, functions like `.encode()` and `.decode()`
 2. Explore/understand data: `.info()`, `.describe()`, `.head()`, `.tail()`, `.shape`
 3. This will give you a sense of what type of dataset you are dealing with
 4. Unique Values check: Use of `.unique()` function

- Data cleaning
 1. Redundant Columns: Use info from Data Exploration steps
 2. Null Value treatment: Decide if you need to drop null values or impute dummy data into them
 3. Checking out the distribution to impute mean, median or mode. Conversion to dummy variables
 4. Check for data consistency to avoid running into issues while merging especially.
Conversion to correct format for string and date type. Example: convert to lowercase

- Aggregation of data.
 - Use `.mean()`, `.median()` type of functions to extract the best average metric for the purpose
 - Mean vs Median example
 - Join Operation to combine dataset from different data frames
 - Example: outer join, inner join, left join, right join
 - Avoiding use of loops while writing code