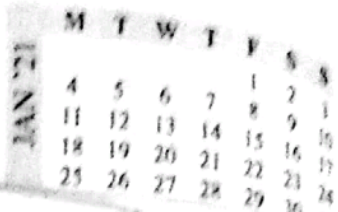
07 January 2021
Thursday
Wk 02 (007-358)



Approval

Loan Prediction System

9.00

10.00 Life cycle of ML project

11.00 1) **Data collection:** Gather relevant data for model training. This may include historical loan data, customer information, credit scores, employment history, and other relevant features. Ensure that data is representative and diverse.

12.00

1.00

2.00

(Loan prediction Problem Dataset (Kaggle.com

3.00

4.00 2) **Data Cleaning**

- Clean the data to handle missing values, outliers, and inconsistencies. This step is crucial for the model's accuracy and generalization

5.00

6.00

7.00 3) **Exploratory Data Analysis (EDA)**

- Conduct EDA to understand the relationships between different variables, identify patterns, and gain insights. Visualization tools can be helpful in this phase

Important Calls	✓	Things to Do	✓	Meetings	✓
	☐		☐		☐
	☐		☐		☐

M T W T F S S
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
FEB '21

i) Feature Engineering

=> Create new features or modify existing ones to improve the model's performance. This might involve transforming variables, creating interaction terms, or encoding categorical variables

ii) Data splitting

=> Split the dataset into training & testing sets. The training set is used to train the model, and the testing sets used to evaluate its performance.

iv) Model selection

=> Choose an appropriate machine learning algorithm for your problem. Common algorithms for loan approval prediction include logistic regression, decision trees, random forests, and support vector machines.

v) Model Training

=> Train the selected model using the training dataset. This involves feeding the algorithm the features and corresponding labels and letting it learn the patterns in the data.

Important Calls ☐ Things to Do ☐ Meetings ☐

09 January 2021
Saturday
WK 02 (009-356)

JAN '21

M	T	W	T	F
4	5	6	7	1
11	12	13	14	8
18	19	20	21	15
25	26	27	28	22
				29

vi) Model Evaluation

9.00 => Evaluate the model's performance on
10.00 the testing dataset using appropriate
metrics such as accuracy, precision,
11.00 recall and f1score

12.00 Data Analysis

Read the CSV file & Obtain the information
1.00 of the dataset such as
df.

2.00 df.info() => Give the datatypes & non null values
of the column

3.00 df.head() => shows first five Rows

4.00 df.shape => shows number of rows & columns

5.00 df.isnull().sum => shows the number of rows
with the null values

Check the Outliers (using boxplot, scatterplot etc)
6.00

fill the null values of numerical datatype
10 Sunday

fill the null values of object type

Check if every column is filled

Important Calls	✓	Things to Do	✓	Meetings	✓
	☐		☐		☐
	☐		☐		☐
	☐		☐		☐

Find Correlation among the columns & visualize with the heatmap

Feature Engineering

- We're ~~merging~~ adding two columns (Applicant Income & Coapplicant income) to calculate the total income.

- After that we can drop Coapplicant Column as It contains many zero values.

Using log function to transform the data

Now, let's convert categorial values into the numerical values

There are two techniques

1. Label Encoding
2. One Hot Encoding

Important Calls	✓	Things to Do	✓	Meetings	✓
	☐		☐		☐
	☐		☐		☐

12 January 2021
Tuesday

Creating Independent and dependent feature

Dependent = Loan status
Independent = Remaining Columns

import the ML models Libs

Algorithms used:

i) Logistic Regression model
ii) Decision Tree Classifier
iii) Random Forest Classifier
iv) KNeighbors Classifier

Train the dataset using these models and test the accuracy of every model

We can use Cross validation technique for more accuracy

Important Calls ✓

Things to Do

✓ Meetings