**#preprocessing**

we applied different techniques such as :

-Feature encoder for transform columns (LoanStatus,BorrowerState,EmploymentStatus) into numeric values

-Drop rows with null values about(27000) for only (LoanRiskScore)->Y column

the data was 107938 rows after this step it becomes 80988 rows.

-Drop columns with more than 60000 null values so we dropped two columns (CreditGrade,TotalProsperPaymentsBilled)

-Fill null values we fill 19 null values in column(EmploymentStatusDuration) with 0

and fill about 6900 value in column (DebtToIncomeRatio) with mean of its column.

-Replace column (IncomeRange) with numeric values for all every range

example we have range 0 to 24000 we replace it with 0.2 ,25000 to 49000 we replace it with 0.4.

-we use normalization to X.

**#Analysis**

A picture containing chart

Description automatically generated

According to this correlation matrix BorrowerAPR ,BorrowerRate columns have high effect on Loan risk score.

**#Regression techniques**

-SVR

-Polynomial Regression.

|  |  |  |
| --- | --- | --- |
| Model | SVR | Polynomial |
| R2 score | 74.2152165214891 | 73.7622454943082 |
| MSE(test) | 1.6085577942432883 | 1.619418339370019 |
| Randomstate | 732 | 150 |
| Training Time | 8 minute 57 sec | 4 sec |

**#Features:- we used all feature as input.**

**#train and test size:**

|  |  |  |
| --- | --- | --- |
| Model | SVR | Polynomial |
| Training size | 0.75 | 0.75 |
| Testing size | 0.25 | 0.25 |

**#Techniques to improve result: -**

-Normalization

