Blueprint

Predicting Prepayment Risk of Mortgage Backed Securities

-Sri Lakshmi Prasanna Koneru

Introduction

- Mortgage-backed securities, called MBS, are bonds secured by home and other real estate loans. They are created when a number of these loans, usually with similar characteristics, are pooled together.
- As the borrowers gradually pay off the underlying mortgage loans, the investors receive payments of interest and principal.
- A large risk factor in MBS lies in the possibility of prepayments.
- Prepayments are payment by borrowers,who pay back a part, or the full amount of the loan earlier than discussed in their mortgage contract.
- Aim of our project is to develop various Machine Learning models that could predict the prepayment risk of mortgage loans by using machine learning techniques like Logistic Regression and Support Vector Machine (SVM) algorithms.

DataSet

- We use Freddie Mac's home loans dataset.
- This dataset contains 291452 rows which represent the number of mortgages/ data points and 28 columns representing different features of the data.
- Let's look at some important features in the dataset.
- **CreditScore:** A number summarising the creditworthiness of the borrower.
- **MIP:** The percentage of loss coverage on the loan - Mortgage Insurance Percentage.
- **Units:** shows whether the mortgage is a 1,2,3 or 4 unit property**.**
- **OCLTV:** Original Combined Loan to Value ratio.
- **DTI:** Debt to Income ratio.
- **OrigUPB:** The unpaid balance of the mortgage on the note date.
- **LTV:** Loan to Value ratio.
- **OrigInterestRate:** The original interest rate as indicated on the mortgage note.

Contd…

- **OrigLoanTerm:** The number of scheduled monthly payments based on the FirstPaymentDate and MaturityDate.
- **NumBorrowers:** The number of borrowers issued on the loan.
- **EverDelinquent:** Has the loan ever been 30 days or more delinquent?
- **MonthsDelinquent:** The number of months the loan has been delinquent.
- **MonthsInRepayment:** The number of months that the loan premium has been paid.
- **FirstPaymentDate:** Date of first payment of interest on the loan.
- **MaturityDate:** Date of expiry of loan payment or date when the loan will be fully paid off.
- **FirstTimeHomebuyer:** Is the purchaser a first time home buyer?
- **MSA:** Marketing Services Agreement.
- **Occupancy:** Type of occupancy in the property mortgaged. O - Owner Occupied, I - Investment Property, S - Second Home.
- **Channel:** Where did the loan get into the notice of borrower. T- Third Party, B - Broker, C - correspondent, R - Retail.

Contd…

- **PPM:** Prepayment Penalty Mortgage. Most of the borrowers are not obligated to pay this penalty.
- **ProductType:** There is only one product type - FRM - Fixed Rate Mortgage.
- **PropertyState:** The state in which the mortgaged property is located. Most of the properties are located in state CA.
- **PropertyType:** Denotes the type of property mortgaged. SF - Single Family, PU - Planned Unit Development, CO - Condominium, MH - Manufactured Home, LH - Lease Hold, CP - Cooperative Property Share.
- **PostalCode:** Records the postal code of the property.
- **LoanSeqNum:** Unique number given to each loan.
- **LoanPurpose:** Indicates whether the mortgage loan is a Cash-out reference mortgage(C), No cash-out Reference Mortgage(N) or a Purchase Mortgage(P).
- **SellerName:** Name of the Seller.
- **ServicerName:** Name of the Servicer.

Steps in this project

1. EDA

Exploratory Data Analysis is the basic step in order to understand the data better. Box plots, distribution plots, scatter plots etc can be used to do this.

2. Data Cleaning and Preprocessing

The data needs to be cleaned - removing outliers and filling missing values. Preprocessing is done to convert the data into consumable format for machine learning models.

3. Model Building and Evaluation

We will be using logistic regression, support vector machine, gaussian discriminant analysis and feed forward neural networks.

4. Deployment

We need to deploy the model on the internet using API's. This can be achieved by using frameworks like Django and Flask and platforms like Heroku, AWS etc.