**MORTGAGE BACKED SECURITIES:**

**PREPAYMENT RATE PREDICTION**

1. **Introduction**

Mortgage-backed securities (MBS) are contractual instruments that grant their holders the rights to the cash flows generated by mortgage loans. However, these investments carry a specific risk known as prepayment risk. Prepayment risk refers to the possibility of investors receiving unscheduled, early principal payments, which occurs when mortgage borrowers have the ability to pay off their mortgages before their scheduled due dates.

When principal is returned earlier than expected, investors face certain disadvantages. They often lose the premium they paid over the face value of the MBS, when prepayment occurs, investors must reinvest at current market interest rates, which are usually substantially lower.

Through our model, we aim to determine the rate of prepayment for a pool of MBS, enabling investors to assess the level of prepayment risk associated with their investments and further help them make informed decisions and prepare for potential declines in payments accordingly.

This model focuses on exogenous configurations and aims to develop an innovative model capable of predicting the single month mortality (SMM) , Cash inflows , Weighted Average Life(WAL) , Weighted Average Maturity(WAM) for a mortgage portfolio based on a set of input features. We have figured out three machine learning algorithms namely Dense neural networks, recurrent neural networks (such as LSTM) and Random Forest which will be employed to achieve this objective.

1. **Motivation**

Working on the complex task of forecasting prepayment rates in Mortgage-Backed Securities (MBS) presents an intellectual challenge that demands expertise in various fields, including finance, economics, and data analysis. This interdisciplinary approach is necessary to develop accurate models and methodologies for predicting the prepayment behavior of MBS, considering the intricate dynamics of the mortgage market. The industry relevance of this endeavor cannot be understated, as it directly impacts stakeholders such as lenders, investors, regulators, and policymakers. By contributing to the understanding and mitigation of risks associated with mortgage financing in MBS, working on this problem plays a crucial role in shaping the stability and functioning of mortgage markets. Moreover, engaging in such challenging research fosters continuous professional growth by facilitating collaboration with finance experts, enabling networking opportunities with industry professionals, and ultimately leading to recognition for valuable contributions in this specialized field.

1. **Approach**

Predicting prepayments in mortgage-backed securities (MBS) involves utilizing the Outstanding Principal Balance to forecast Single Month Mortality (SMM) , cash inflows and Weighted Average Maturity (WAM). SMM calculates the monthly rate of loan prepayment, while WAM estimates the average maturity of mortgage loans whereas an entirely different model was used to calculate WAL utilizing months to complete payment of principal as output variable. By analyzing historical data, developing models, and considering market conditions, investors can make informed decisions and manage prepayment risk in MBS investments.

**3.1 Calculating Prepayment**

The SMM can be used to calculate a dollar amount for the monthly principal prepayment.

We would predict prepaid principal for individual mortgages using our model and utilize CPR model to calculate prepayment rates.

**3.2 Calculating Cash Inflows**

Cash inflows in mortgage-backed securities are payments from borrowers that provide income to investors. They come in the form of monthly principal and interest payments, crucial for evaluating MBS investment potential.

It is calculated using the following formula.

**3.3 Calculating Weighted Average Life**

Weighted Average Life (WAL) is a metric used to compute time in months before which the entire principal of the pool will be prepaid weighed by the principal amount of individual loans.

It is calculated using the following formula.

**3.4 Calculating Weighted Average Maturity**

Weighted Average Maturity (WAM) in MBS refers to the average time it takes for mortgage loan principal payments to be received by investors. It considers the remaining time to maturity and outstanding principal balance of each loan in the MBS pool.

1. **Scope**

The scope of our MBS (Mortgage-Backed Securities) prepayment rate machine learning project can be expanded by including additional attributes such as housing price index, unemployment rate, divorce rate, and inflation. These additional attributes can provide valuable insights into the factors influencing prepayment rates.

Housing Price Index: The housing price index can be used as a proxy for the overall health of the real estate market.

Unemployment Rate: Unemployment rate is a key economic indicator that reflects the general employment situation.

Divorce Rate: Divorce rate can be an interesting attribute to consider as it may reflect changes in household composition and financial circumstances.

Inflation: Inflation is a measure of the general increase in prices over time. It can affect interest rates and borrowers' purchasing power, which in turn can impact prepayment rates.

**4.1 Prepayment Scenarios**

Scenario 1: Early Refinancing

A homeowner with a mortgage-backed security (MBS) decides to refinance their mortgage before the maturity date. This occurs when interest rates decrease, allowing the homeowner to obtain a new loan with a lower interest rate.

Scenario 2: Home Sale

A homeowner decides to sell their property before the mortgage reaches its maturity date. When the property is sold, the proceeds from the sale are used to repay the mortgage associated with the MBS.

Scenario 3: Loan Default

In some cases, a homeowner may default on their mortgage payments, resulting in foreclosure proceedings. The foreclosure process allows the lender to sell the property to recover the outstanding loan amount. When a property is foreclosed and sold, the proceeds are used to repay the mortgage tied to the MBS.

Scenario 4: Bulk Prepayments

In certain situations, a group of borrowers within a pool of mortgages may make prepayments simultaneously. This can happen if interest rates drop significantly, prompting many borrowers to refinance their loans.

Scenario 5: Extra Principal Payments

Homeowners occasionally make extra principal payments on their mortgages. These additional payments reduce the outstanding loan balance, resulting in early repayment of the mortgage and a prepayment of the MBS.

**4.2 PPS & PSA**

Prepayment speed refers to the rate at which borrowers repay their mortgages early, either through refinancing, home sales, or other means. In the context of Mortgage-Backed Securities (MBS), the Public Securities Association (PSA) provides a standardized method to estimate prepayment speeds. The PSA benchmark expresses prepayment speeds as a percentage of the Single Monthly Mortality (SMM) rate, which represents the proportion of outstanding mortgage balances that are prepaid each month. The PSA assumes a gradual increase in prepayment speeds, starting at 0.2% in the first month and rising by 0.2% each month until reaching 6% in the 30th month. After the 30th month, a constant prepayment speed of 6% is assumed. These standardized calculations help MBS investors assess and model the expected cash flows and durations of MBS investments.

**4.3 Future Enhancements**

* **Larger Dataset:** Utilize a larger dataset for training to get better results.
* **CPR PSA Calculation:** Improve accuracy by adjusting prepayment projections for 12 months.
* **Extension to IO and PO loans:** Include interest-only and principal-only loans in the analysis.
* **Predict Prepayment and Default:** Enhance models to predict both prepayment and default behavior.
* **Scenario Analysis and Stress Testing:** Incorporate scenario analysis and stress testing for risk assessment and contingency planning.

1. **DataSet**

**5.1. Fannie Mae Single Loan Level dataset**

This involves : Loan-level origination, monthly loan performance, and actual loss data on fully amortizing fixed-rate Single-Family mortgages that Fannie Mae acquired with origination dates from 2000 to 2020. This population closely resembles mortgages that are eligible for the Single-Family Credit Risk Transfer (“CRT”) transactions.

**5.2. Dataset Attributes Significant to Problem**

1. ORIGINAL UPB - The unpaid principal balance of the mortgage on the note date.

2. Borrower Credit Score at Origination : A number, prepared by third parties, summarizing the borrower’s creditworthiness, which may be indicative of the likelihood that the borrower will timely repay future obligations.

3. FIRST TIME HOMEBUYER FLAG - Indicates whether the Borrower, or one of a group of Borrowers, is an individual who (1) is purchasing the mortgaged property, (2) will reside in the mortgaged property as a primary residence and (3) had no ownership interest (sole or joint) in a residential property during the three-year period preceding the date of the purchase of the mortgaged property.

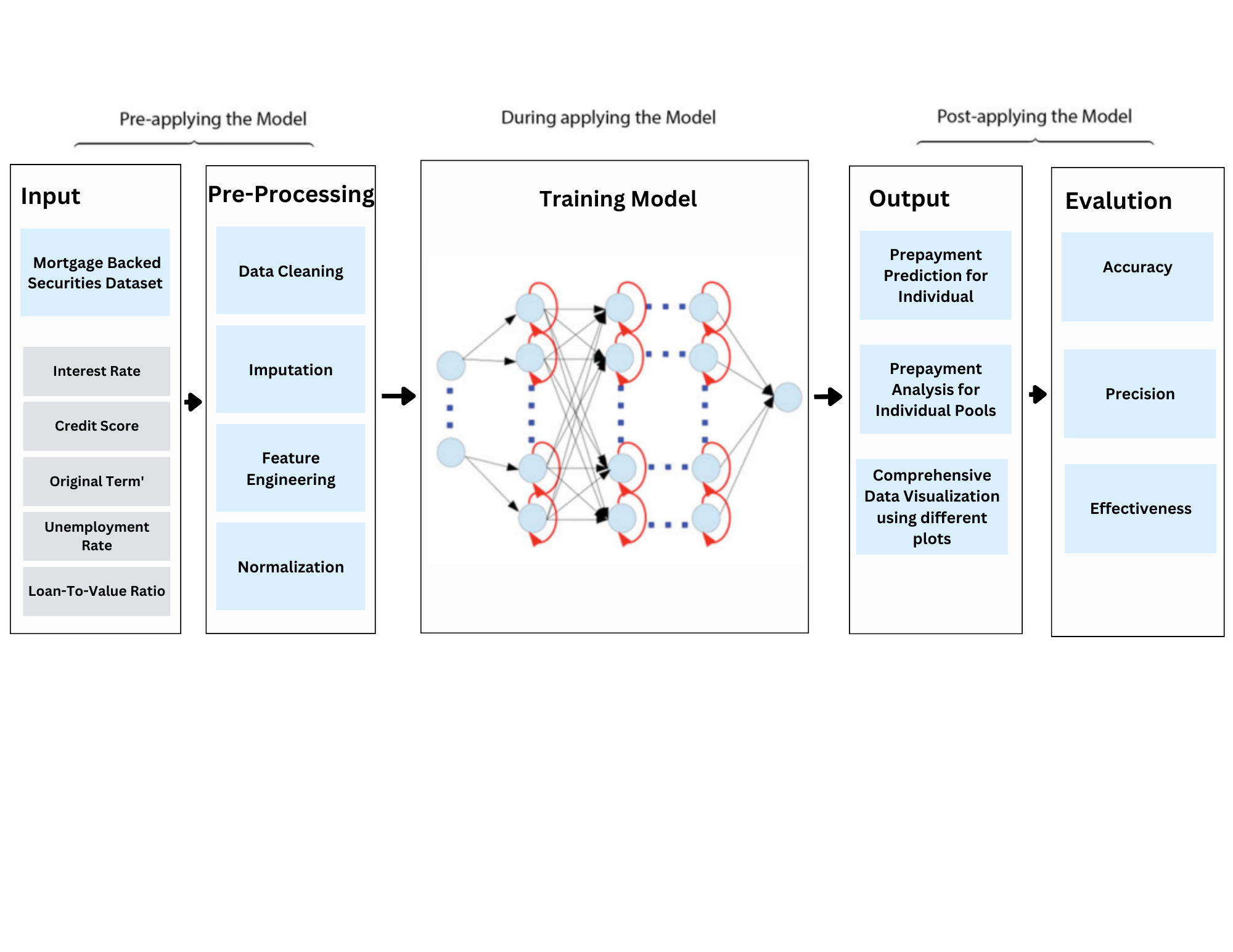
4. Remaining Months TO Maturity - The month in which the final monthly payment on the mortgage is scheduled to be made as stated on the original mortgage note.

5. ORIGINAL LOAN-TO-VALUE (LTV) - In the case of a purchase mortgage loan, the ratio obtained by dividing the original mortgage loan amount on the note date by the lesser of the mortgaged property’s appraised value on the note date or its purchase price.

6. ORIGINAL DEBT-TO-INCOME (DTI) RATIO - Disclosure of the debt to income ratio is based on (1) the sum of the borrower's monthly debt payments, including monthly housing expenses that incorporate the mortgage payment the borrower is making at the time of the delivery of the mortgage loan to Freddie Mac, divided by (2) the total monthly income used to underwrite the loan as of the date of the origination of the such loan.

7. METROPOLITAN STATISTICAL AREA (MSA) OR METROPOLITAN DIVISION - This disclosure will be based on the designation of the Metropolitan Statistical Area or Metropolitan Division as of the date of issuance.

**5.3 Data flow**

****

**Correlation matrix:**

A correlation matrix is a statistical tool used to examine the relationship between multiple variables in a dataset. It provides valuable insights into the strength and direction of associations between variables, allowing us to understand the interdependencies and patterns within the data. The significance of a correlation matrix lies in several key aspects:

1. Understanding Relationships

2. Multivariate Analysis

3. Variable Selection

4. Data Quality Assurance

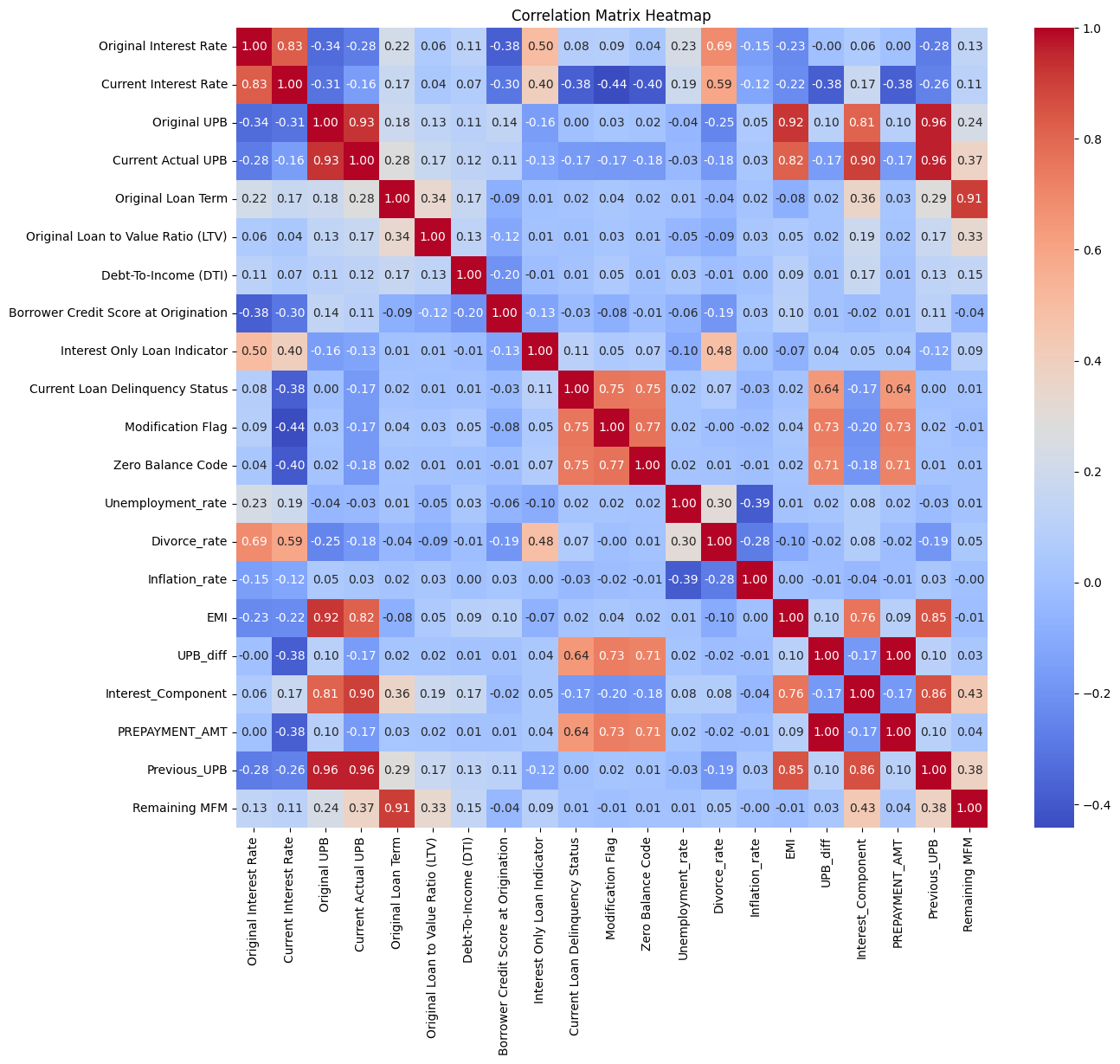
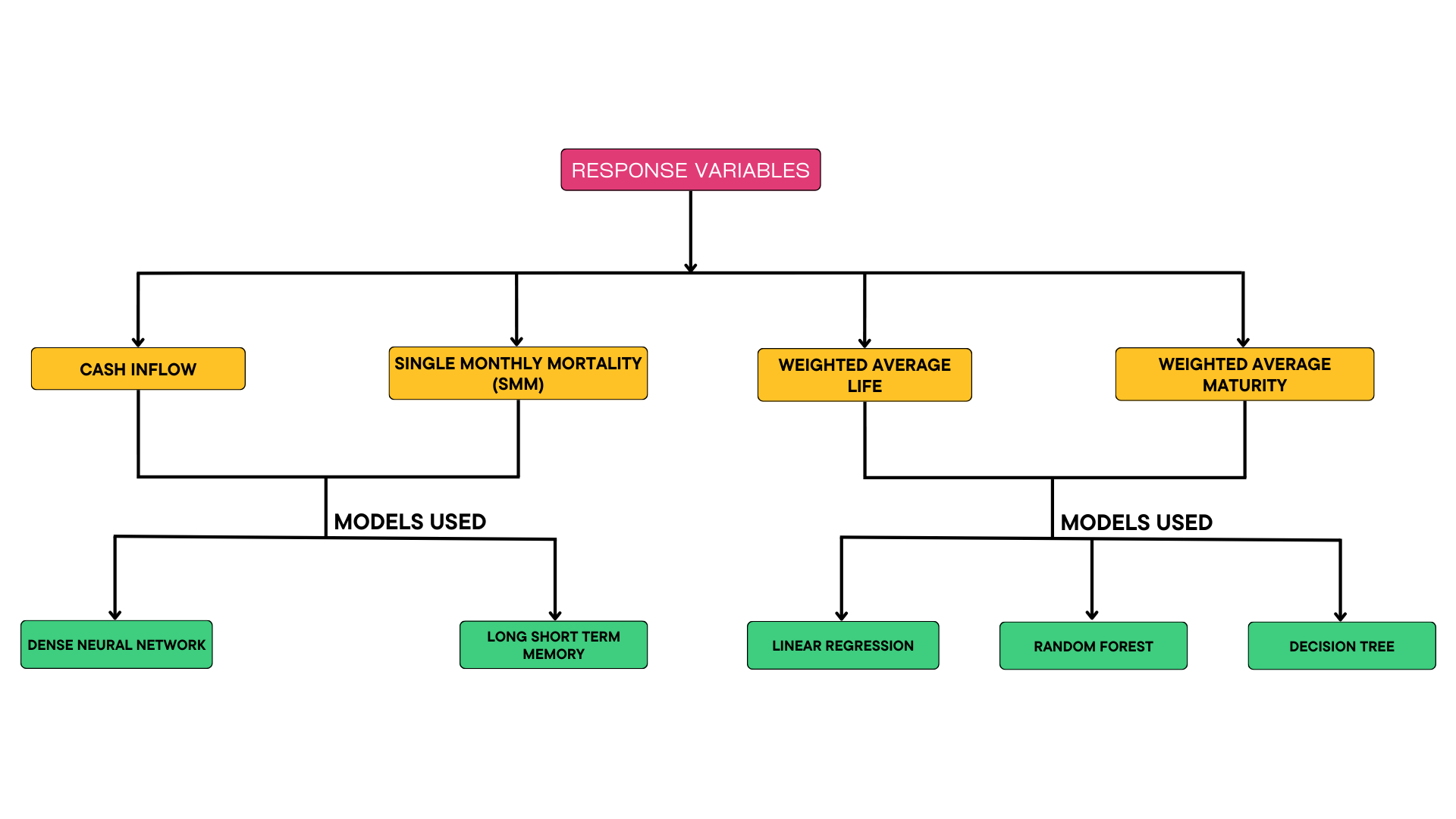


Fig.1. Correlation matrix of dataset

From the above matrix we can observe many correlations within the attributes of the datasets. Below we have stated few of the important relations:

1. Relationship between Current Actual UPB and EMI: There is a strong positive correlation (0.818) between the Equated Monthly Installment (EMI) and the Current Actual UPB, indicating that as the EMI increases, the current unpaid balance tends to increase.
2. Relationship between Divorce rate and Original Interest Rate: There is a positive correlation (0.685) between the divorce rate and the original interest rate. Higher divorce rates are associated with higher original interest rates.
3. Relationship between Divorce rate and Current Interest Rate: Similarly, there is a positive correlation (0.592) between the divorce rate and the current interest rate. Higher divorce rates are associated with higher current interest rates.
4. Relationship between PREPAYMENT\_AMT and Zero Balance Code: There is a strong positive correlation (0.713205) between the prepayment amount and the zero balance code. This indicates that higher prepayment amounts are associated with a higher likelihood of loans reaching a zero balance.
5. Relationship between EMI and Interest\_Component: There is a strong positive correlation (0.758490) between the EMI and the interest component of the payment. This suggests that higher EMIs are associated with higher interest components.
6. **Methodology and Comparison between different Models**

We employed a combination of Deep Neural Network (DNN) and Long Short-Term Memory (LSTM) models to calculate Single Monthly Mortality (SMM) and Cash Inflow. Additionally, for Weighted Average Life (WAL) and Weighted Average Maturity (WAM) estimation, we utilized Random Forest, Decision Tree, and Linear Regression models. After training the data on these models, we observed that DNN yielded the most accurate results for SMM and Cash Inflow calculation, while the Decision Tree model demonstrated superior performance for WAL and WAM calculation. These findings highlight the effectiveness of these models in predicting prepayments and assessing the average life and maturity of mortgage-backed securities.



**6.1 Random Forest**

Random Forest is a suitable model for MBS prepayment prediction due to its ability to handle tabular data and capture complex interactions between loan attributes. For instance, consider a scenario where we have attributes such as loan age, current actual UPB, and current interest rate. Random Forest can effectively analyze these attributes and identify their influence on prepayment rates. For example, the model may discover that loans with higher loan ages and lower current actual UPB tend to have higher prepayment rates, indicating that borrowers are more likely to pay off their loans as they age and the principal balance decreases. Additionally, the model may reveal that loans with lower interest rates are associated with higher prepayment rates, as borrowers may seek to refinance at lower rates. By considering these attribute interactions, Random Forest can provide valuable insights into the factors driving prepayment behavior in MBS portfolios.

By leveraging these time series factors alongside other loan attributes, Random Forest can provide comprehensive insights into the drivers of prepayment behavior in MBS portfolios over time.

**6.2 Decision Tree**

A decision tree is a predictive model that uses a tree-like structure to make decisions based on features and outcomes. The tree consists of a root node, internal nodes, branches, and leaf nodes. The root node represents the initial decision point, while internal nodes split the data based on specific conditions. Branches represent outcomes based on different feature values, leading to leaf nodes that indicate predicted outcomes. Decision trees are interpretable, providing insights into the factors driving prepayment behavior in MBS.

**6.3 Linear Regression**

Linear regression is a statistical technique that models the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship and aims to find the best-fit line that minimizes the difference between predicted and actual values. The coefficients are estimated using ordinary least squares (OLS) to minimize the sum of squared differences. It is commonly used for prediction and analysis in different fields.

**6.4 Dense Neural network**

DNN models can capture complicated feature interactions and non-linear correlations, making them suited for the intricate relationships found in MBS prepayment data. In MBS datasets, for example, characteristics such as borrower credit score, loan-to-value ratio, interest rate, loan amount, and loan age interact in complex ways, and DNN models can successfully capture these relationships. DNN models can also automatically extract important features, minimizing the requirement for manual feature engineering. This is especially beneficial in MBS data, which frequently comprises high-dimensional and diversified loan information. DNN models are scalable, allowing them to handle large-scale MBS datasets easily. They generalize effectively to fresh data, allowing for accurate projections on new loans or under various market conditions.

DNN architectures' flexibility allows for modification to the specific aspects of MBS prepayment prediction. For example, to improve the model's performance, activation functions, and regularization techniques. DNN models can also use transfer learning to improve performance by using pre-trained models on comparable tasks or datasets.

**6.5 Long Short-Term Memory**

LSTM models can capture sequential dependencies, learn long-term interactions, and manage variable-length input sequences, which makes them useful for forecasting Mortgage-Backed Securities (MBS) prepayment rates. Think about a situation where the loan features include the borrower's prior payment patterns, the length of the loan, changes in interest rates, and economic conditions. The borrower's payment patterns over time can be accurately captured by an LSTM model, which can also learn how the borrower's payment habits affect judgements about future prepayments. The model may identify trends like a borrower's rising delinquencies or constant on-time payments, which can affect prepayment rates, by taking into account the sequential nature of the data.

LSTM models are specifically designed for processing sequential and time-series data, making them ideal for analyzing the historical patterns and dependencies within the MBS dataset. Prepayment rates in MBS can be accurately modeled using LSTM models by capturing the sequential pattern of the monthly reporting periods and loan parameters.

Additionally, LSTM models are excellent at capturing continuous interactions between characteristics. Prepayment rates, for instance, may be impacted by a mix of loan duration, interest rate changes, and economic conditions. An LSTM model can be used to capture the intricate linkages and determine how they affect changes in prepayment rates by looking at historical trends and patterns in these parameters.

1. **Evaluation Metrics**

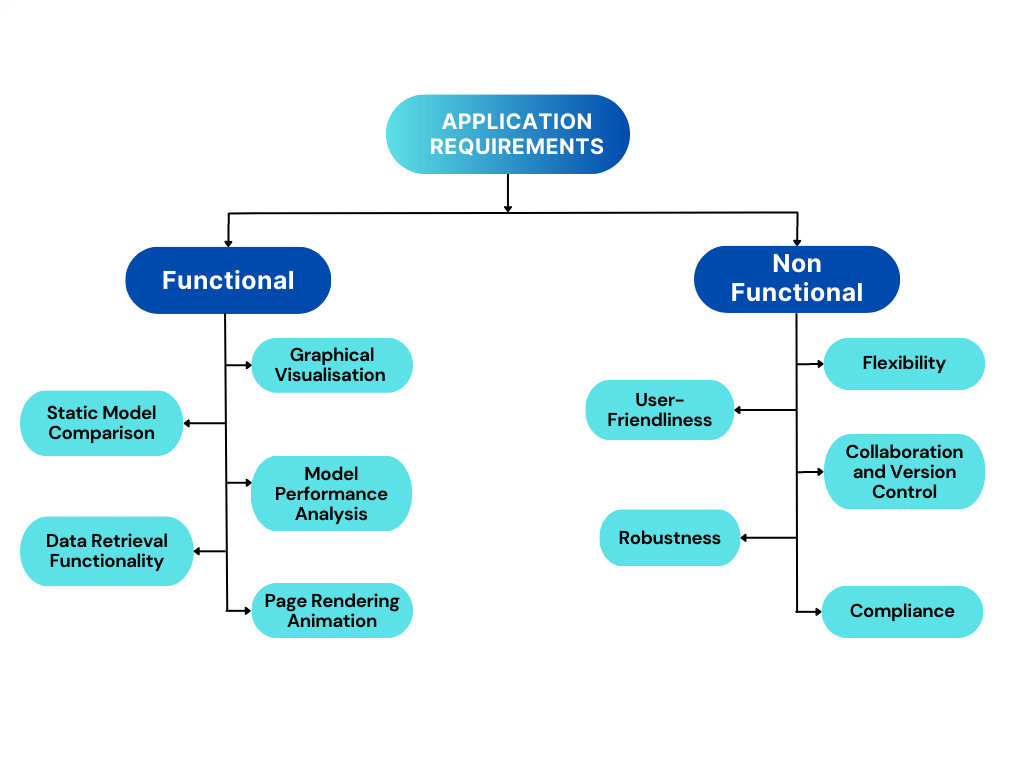
* Models Used for SMM and Cash Inflow Calculations

|  | DNN | LSTM |
| --- | --- | --- |
| R2 Score | 0.9991 | 0.9969 |
| Root Mean Square Error | 3255.8909 | 5823.015 |
| Mean Absolute Error | 281.429 | 382.637 |

* Models Used for WAL and WAM Calculations

|  | Decision Tree | Random Forest |
| --- | --- | --- |
| R2 Score | 93.0920 | 95.2086 |
| Mean Square Error | 118.4987 | 79.8623 |
| Mean Absolute Error | 3.1961 | 2.5098 |

1. **Design Requirements**



**8.1 Functional Requirements:**

System should integrate and process various data sources, including historical mortgage data, economic indicators, interest rate data, and borrower-specific information. It should apply appropriate models and algorithms to calculate prepayment rates based on the input parameters and historical data.

Additionally, the system should determine the prepayment rate for different pools. It should generate overall prepayment rate estimates for the housing loan market, considering a wide range of parameters and data sources. The system should provide visualizations of prepayment rates for effective analysis and interpretation. Furthermore, it should enable historical analysis of prepayment rates, allowing users to explore trends, patterns, and correlations with other variables over time.

**8.2 Non-functional Requirements:**

The non-functional requirements include compliance with relevant legal and regulatory requirements, such as data privacy laws and financial industry regulations. The system should be flexible enough to incorporate new data sources or adjust the model as market conditions or industry standards change. It should have an intuitive and user-friendly interface, enabling users to easily input data, interpret results, and interact with the system effectively. The system should be robust and resilient, capable of handling data anomalies, missing values, or unexpected scenarios, while maintaining consistent performance and accurate predictions. Additionally, it should support collaboration among multiple users or teams, providing features for version control, code sharing, and collaboration on model development and evaluation.

1. **Front-End Design**
2. **Deployment Plan**

We plan to deploy the model on a hosting platform like Heroku or directly on the cloud servers like AWS by setting up the runtime environment and configuring necessary libraries and packages, the code can be uploaded on the platform.

1. **Testing And Quality Assurance**

General outline of the testing plan for our project is as follows :

**Testing**: The model's accuracy and correctness are being evaluated through testing with real-world unseen data.

**Data Quality Testing**: Validate the quality and integrity of data. Check for missing values, outliers, and inconsistencies. Ensure that our data is representative of the problem we're trying to solve and that it covers a wide range of scenarios.

**Model Evaluation**: Evaluation of trained machine learning models using appropriate evaluation metrics such as accuracy, precision, recall, F1 score, or area under the receiver operating characteristic curve (AUC-ROC).

**Hyperparameter Tuning**: Optimizing model's hyperparameters to improve its performance. Use techniques like grid search, random search, or Bayesian optimization to find the best combination of hyperparameters.

**Model Robustness Testing**: Testing model's performance under different scenarios. Check how well it handles missing or noisy data, unexpected inputs.

1. **Tech Stack**

AI/ML:

* Numpy
* Pandas
* Keras
* Matplotlib
* Seaborn
* Scikit Learn
* Tensor Flow

Front-End Development:

* HTML
* CSS
* JavaScript

Back-End Development:

* Flask

1. **APPENDIX**

Appendix: Loan Types, Loan Performance, and Loan Status in Freddie Mac Single Loan-Level Dataset

1. Loan Types:

1.1 Fully Amortizing 15-, 20-, 30-, and 40-year Fixed-Rate Mortgages:

- Mortgages with a fixed interest rate and a set repayment term.

- Regular monthly payments including both principal and interest.

- Loan fully repaid by the end of the term.

1.2 Mortgages Categorized as Having Verified or Waived Documentation:

- Different levels of documentation required during the mortgage application process.

- "Full documentation" involves comprehensive verification of income, assets, and employment.

- Some mortgages offer "waived documentation" or "stated income" loans with less extensive verification.

1.3 Relief Refinance Mortgages:

- Refinance loans designed to provide potential savings and financial relief.

- Allow borrowers to obtain more favorable terms, such as lower interest rates or reduced monthly payments.

- Aimed at assisting homeowners facing financial difficulties or taking advantage of improved market conditions.

1.4 Home Possible Mortgages:

- Mortgage program aiding low to moderate-income borrowers with affordable homeownership options.

- Flexible down payment requirements, reduced mortgage insurance costs, and lenient eligibility criteria.

- Offers options for both fixed-rate and adjustable-rate mortgages.

2. Loan Performance Information:

2.1 Prepaid or Matured (Voluntary Payoff):

- Borrower chooses to pay off the mortgage loan in full before the scheduled maturity date.

- Voluntary action usually involving a lump sum payment.

2.2 Third Party Sale:

- Property securing the mortgage loan sold to a buyer who is not the original borrower.

- Proceeds from the sale used to pay off the existing mortgage loan.

- New owner assumes responsibility for the mortgage or secures a new loan.

2.3 Short Sale or Charge Off:

- Lender agrees to accept a sale price for the property less than the outstanding mortgage balance.

- Occurs when the borrower is unable to make mortgage payments and owes more than the property's value.

2.4 Repurchase prior to Property Disposition:

- Lenders repurchases the mortgage loan from an investor or securitization trust before property sale.

- Done for various reasons, such as dispute resolution, delinquency resolution, or mitigating losses.

2.5 REO Disposition:

- REO refers to properties owned by the lender through foreclosure.

- Lender sells the property to recover the outstanding debt.

- REO properties sold through traditional real estate channels or auctions.

2.6 Note Sale:

- Sale of the mortgage loan, including the promissory note and associated rights, to another party.

- May occur to free up capital, reduce risk exposure, or transfer servicing rights.

2.7 Reperforming Loan Sale:

- Reperforming loan is a previously delinquent loan that has become current again.

- Sale involves transferring these loans from the original lender or investor to another party.