**Project Outline**

**Title:**

**Cross-Country Prediction of Overnight Interest Rate Changes Using Demographic and Macro-Structural Features**

**⸻**

**1. Introduction**

**• Background & Motivation**

**• Importance of overnight rates for monetary policy and financial markets**

**• Most existing works focus on single-country models; cross-country prediction is rare but crucial in a globalized economy**

**• Research Gap**

**• Can demographic and structural macro features help predict overnight rate changes across major economies?**

**• Contributions**

**• Constructing a cross-country dataset**

**• Designing a machine learning pipeline leveraging demographic & macro features**

**• Benchmarking with traditional models**

**⸻**

**2. Related Work**

**• Single-country overnight rate prediction: ARIMA, VAR, Random Forests, Neural Networks**

**• Cross-country macroeconomic forecasting (GDP, inflation, etc.)**

**• Recent ML advances in temporal and multi-variate forecasting (TFT, LSTM, XGBoost, etc.)**

**• (Highlight gap: lack of cross-country overnight rate ML studies)**

**⸻**

**3. Problem Definition**

**• Prediction targets**

**• (a) Will the overnight rate change within the next N days? (Classification)**

**• (b) How much will it change? (Regression)**

**• Input features**

**• Demographic (population, growth rate, gender ratio, birth cohorts, etc.)**

**• Macro-structural (GDP, GDP growth, inflation, unemployment, fiscal balance, etc.)**

**• Historical overnight rate (lags, deltas)**

**• Country fixed effects / region dummies**

**• Evaluation Metrics**

**• For (a): Accuracy, Precision, Recall, F1**

**• For (b): MAE, RMSE, R²**

**⸻**

**4. Data Collection and Preprocessing**

**• Target countries: G20 or top 20 GDP economies**

**• Data sources: Central banks, IMF, World Bank, UN, CEIC, Bloomberg, OECD**

**• Feature engineering:**

**• Compute derivatives/growth rates**

**• Normalize across countries**

**• Align time series**

**• Handle missing data (imputation/interpolation)**

**• Generate lags, rolling windows**

**• Data split strategy: Time-based split (train/val/test chronologically)**

**⸻**

**5. Methods**

**• Baseline Models**

**• Univariate time series (AR, ARIMA, random walk)**

**• Country-wise regression**

**• Machine Learning Models**

**• Logistic/Linear Regression**

**• Random Forest, Gradient Boosted Trees (XGBoost/LightGBM)**

**• LSTM/Temporal Fusion Transformer (for multi-country, multi-variate time series)**

**• Multi-task or transfer learning (optional, for joint country learning)**

**• Feature selection & importance analysis**

**• Hyperparameter tuning: Cross-validation, grid/random search**

**⸻**

**6. Experiments**

**• Experiment design**

**• Benchmark: Baseline vs. ML**

**• Single-country vs. cross-country joint models**

**• Ablation: impact of demographic vs. economic vs. historical features**

**• Results**

**• Quantitative results (tables/plots of metrics)**

**• Feature importance visualization**

**• Case studies: Model performance during global events (e.g., COVID-19 shock)**

**• Error analysis**

**• Country-specific vs. global patterns**

**• Where/why the model fails**

**⸻**

**7. Discussion**

**• Interpretation**

**• What features are most predictive across countries?**

**• Are demographic trends globally significant, or country-specific?**

**• Practical implications for policy and financial market forecasting**

**• Limitations**

**• Data availability/quality**

**• Structural breaks, country heterogeneity**

**• Model assumptions**

**⸻**

**8. Conclusion**

**• Summary of findings**

**• Contributions and innovation**

**• Directions for future work (e.g., higher-frequency data, more countries, alternative features)**

**⸻**

**9. References**

**• Cited papers, data sources, model docs**

**⸻**

**10. Appendix**

**• Data pipeline scripts, extra figures, reproducibility notes**