Timeline pre-thesis 2025-2026  
Week 1:

- Download datasets (Kaggle)

- Write data card (source, variable, target)

- Do preliminary EDA

- Read related papers to understand the problem

Week 2:

- Data cleaning (missing, outlier)

- Categorical encoding

- Imbalance handling

- Deep EDA

Week 3:

- Build baseline models (Majority, Logistic Regression, Decision Tree)

- Set up Cross-validation

- Write evaluation function (Accuracy, Precision, Recall, F1, ROC-AUC)

Week 4:

- Train advanced models (RF, XGBoost, SVM, Neural Net)

- Save CV results on tables

- Preliminary comparison of classical vs. advanced

Week 5:

- Hyperparameter tuning (Optuna/RandomizedSearchCV)

- Try class\_weight/SMOTE in pipeline

- Evaluate the optimal model

Week 6:

- Calculate Feature Importance (RF, XGBoost)

- Use SHAP to explain the model

- Draw graphs (SHAP beeswarm, bar, PDP/ICE)

- Write RQ1 findings

Week 7:

- Compare classical vs. advanced models (RQ4)

- Test statistics (Wilcoxon)

- Write RQ2–RQ4 findings

- (Optional) build mock demo early warning (RQ5)

Week 8:

Slide + report