ClaimWise

SMARTER INSURANCE
THROUGH RISK REDUCTION

SHARK
TANK

CLAIMWISE TEAM

Weiqi Huang

Shahriar Rahman

Mansour Almubaraki

Fangdi (Flora) Zhai

Carla Francois

Romina Fareghbal

THE PROBLEM

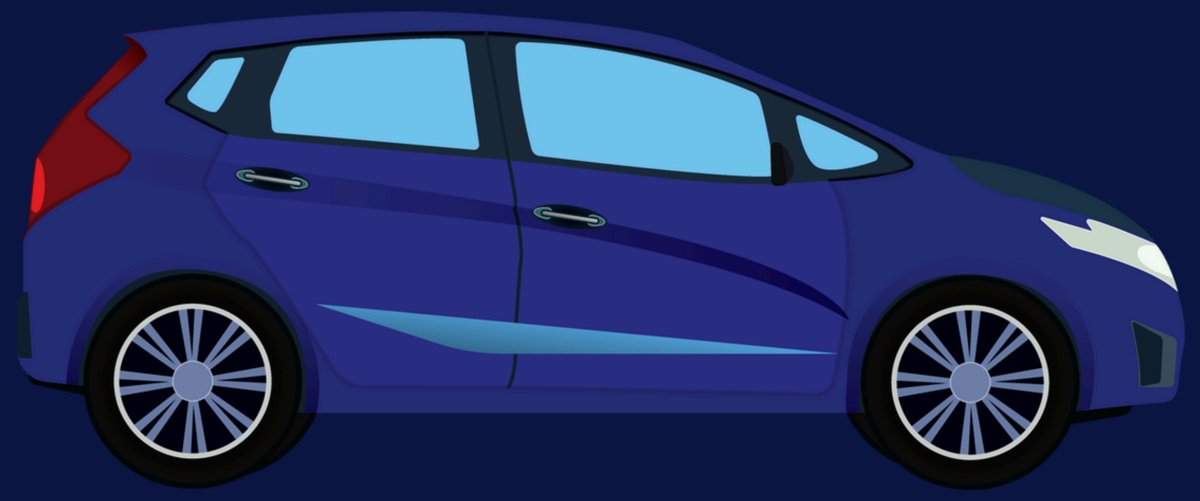
Insurers are losing money due to mispriced policies and surprise claim.

- Overpricing drives away low-risk customers.
- Underpricing pulls in high-risk drivers.
- Hidden risks and fraud slip through.

DUALSOLUTION

Data transformation:

Engineered features including vehicle age, driver age, driving experience, policy duration, and time since last renewal



Phase 1: Predicting Financial Risk

- Predict Loss Cost (LC) and

Historically Adjusted Loss Cost (HALC)

Phase 2: Predicting Behavioral Risk

- Predict who is likely to file a claim

PREDICTING LOSS COST & HALC

- **Generalized Linear Model (GLM) using the Tweedie distribution.**

  LC RMSE: 291.00

  HALC RMSE: 573.65

- **Random Forest, XGBoost, Gradient Boosting, LightGBM, Neural Network.**

	Model	LC MSE	LC RMSE	HALC MSE	HALC RMSE
0	GLM (Tweedie)	84680.790000	291.000000	329070.640000	573.650000
1	Random Forest	88761.470000	297.930000	339787.000000	582.910000
2	Gradient Boosting	83610.060000	289.150000	324290.070000	569.460000
3	LightGBM	82999.900000	288.100000	323034.690000	568.360000
4	XGBoost	83034.960000	288.160000	324819.540000	569.930000
5	Neural Network	85931.460000	293.140000	324680.990000	569.810000

PREDICTING LOSS COST & HALC

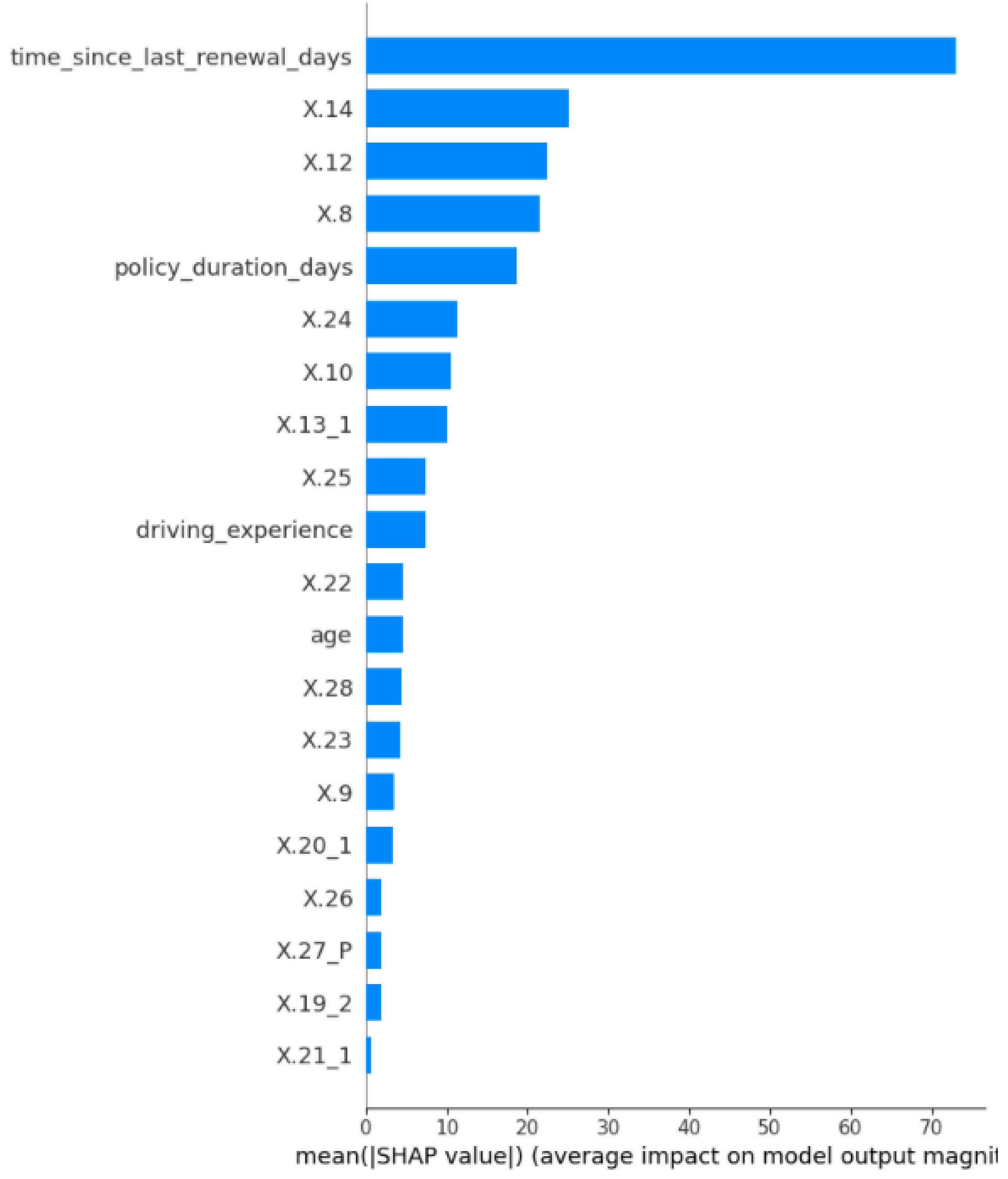
- **Cross Validation - 5-fold**

	Model	LC MSE	LC RMSE	HALC MSE	HALC RMSE
3	LightGBM	80876.470000	284.190000	301875.160000	548.740000
2	Gradient Boosting	82225.260000	286.560000	302763.040000	549.490000
0	GLM (Tweedie)	83252.040000	288.270000	308180.490000	554.360000
5	Neural Network	85930.290000	292.950000	310842.530000	556.820000
1	Random Forest	87007.950000	294.840000	322730.070000	567.610000
4	XGBoost	88265.930000	296.920000	332165.420000	575.890000

- **Best Model - LightGBM**
    - Captures complex patterns through boosting
    - Handles large, sparse data efficiently
    - Fast training and prediction
- **LightGBM Hyperparameter Tuning**
    - LC RMSE: 283.37
    - HALC RMSE : 545.27

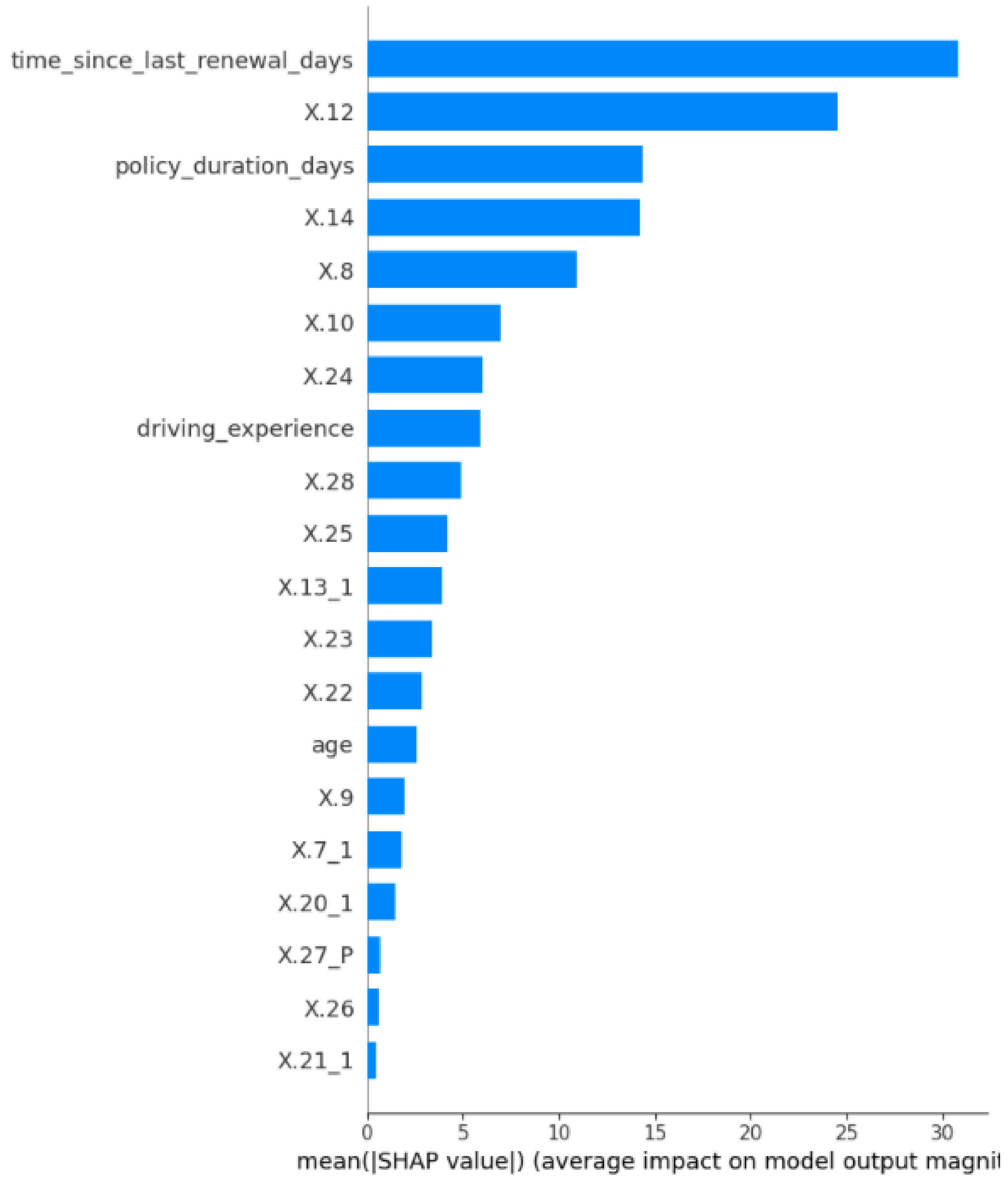
WHAT DRIVES PREDICTIONS?

SHAP BAR CHARTS



HALC

- **SHAP to interpret LightGBM predictions**
- **HALC and LC top 5 features: time _since_last_renewal, policy_duration, X.14, X.12, and X.8**



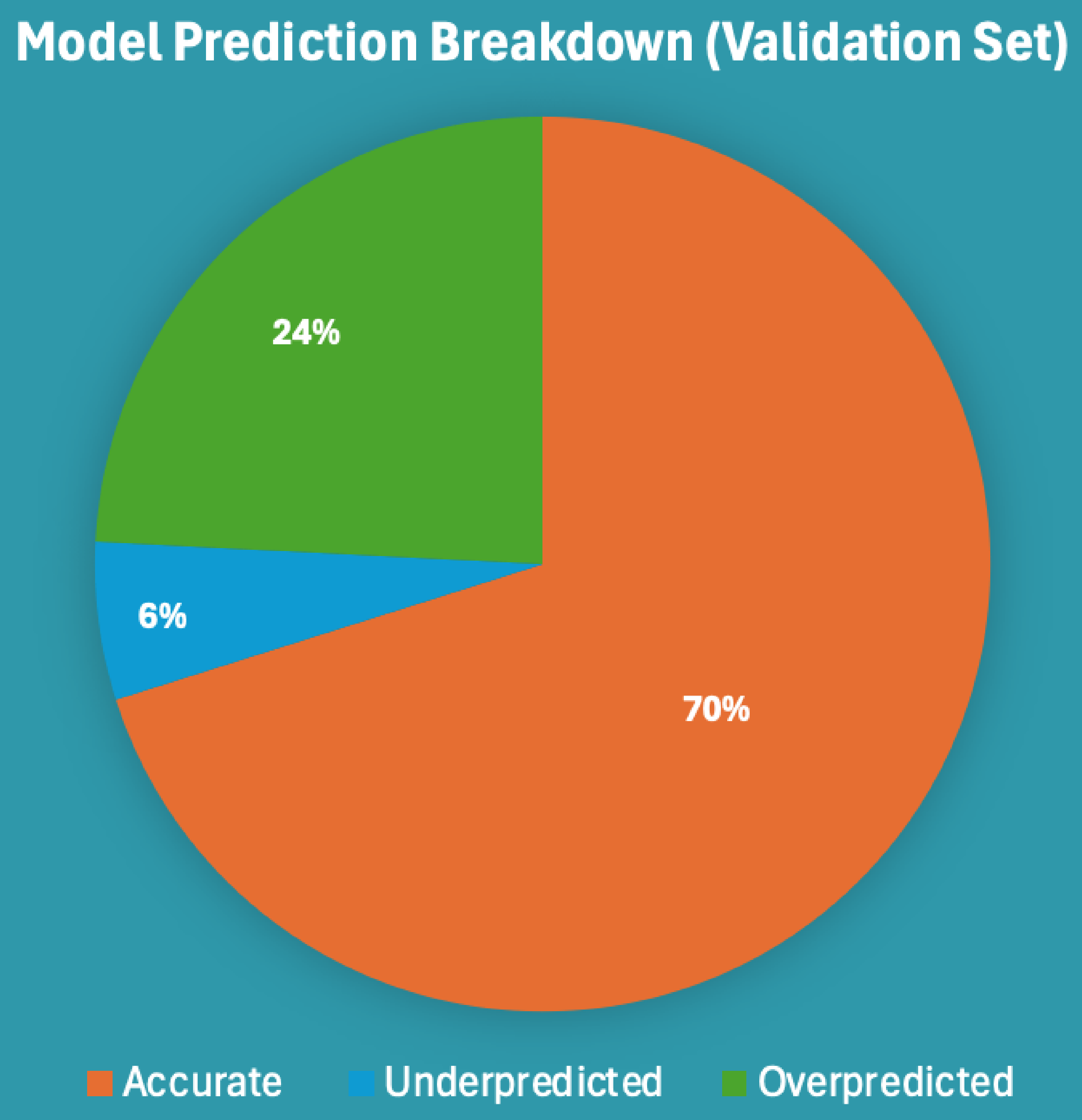
LC

X.8 Total number of years that the insured has been associated with the insurance entity.
X.12 Number of policies canceled or terminated for nonpayment in the currentyear.
X.14 Net premium amount associated with the policy during the current year.

MODEL PREDICTIONS IN ACTION
RISK PROFILES + ACCURACY

```
========================================
Risk Profile – Policyholder #10
----------------------------------------
Predicted HALC:  $17.91
Actual HALC:     $0.00
----------------------------------------
Top Feature Drivers:
                    Feature  Value  SHAP Impact
time_since_last_renewal_days  2557   -95.368409
                       X.12     0    17.411995
                        X.8    13   -10.041212
========================================
```

```
========================================
Risk Profile – Policyholder #25
----------------------------------------
Predicted HALC:  $269.29
Actual HALC:     $0.00
----------------------------------------
Top Feature Drivers:
                    Feature  Value  SHAP Impact
                       X.26     6    88.446288
time_since_last_renewal_days     0    76.314466
                        X.8     1   -60.169905
========================================
```



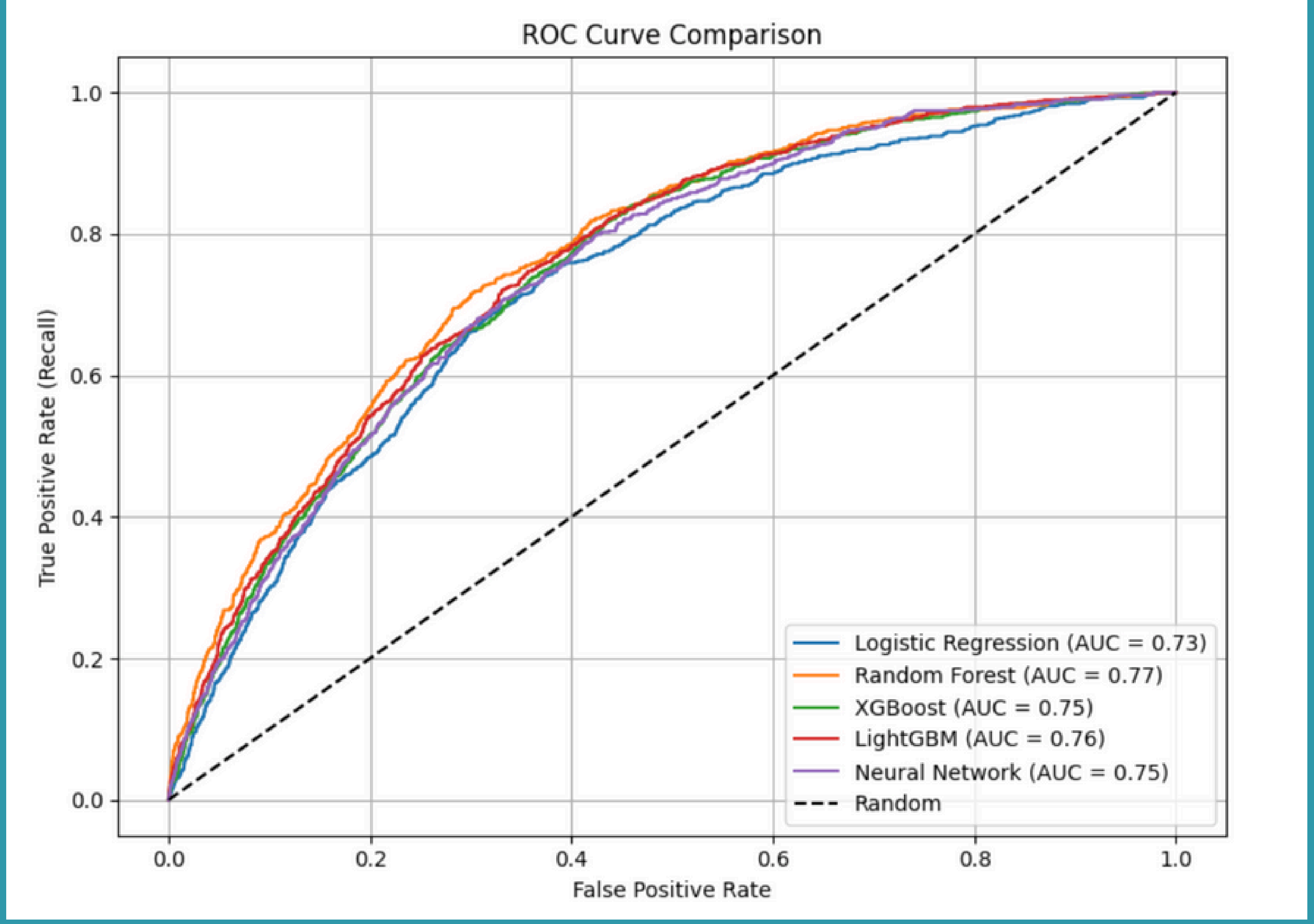
PREDICTING CLAIM STATUS (CS)

- Adverse selection can lead to financial losses for insurance companies.
- Our goal is to predict which policyholders are likely to file a claim.
- Accurate predictions help improve:

  Pricing strategies, Risk segmentation, Portfolio balance
- Machine learning can reduce uncertainty in the underwriting process.

MODEL BUILDING PATH

1. **Model Selection**
   - Validation approach: avoid underfitting or overfitting
   - Use AUC and Accuracy Rate to select the best model
2. **Variable Selection**
   - Apply SHAP, LASSO, RIDGE to select variables
3. **Hyper parameter tuning**
   - Apply k fold cross validation to tune the best model
4. **Final Model**
   - Random Forest Classifier
5. **Future optimization plan - focus on Recall Rate**
   - For claim status prediction model, prioritizing recall ensures high-risk individuals are appropriately flagged and not systematically missed




```python
# Create Random Forest model
model = RandomForestClassifier(
    n_estimators=300,
    min_samples_leaf=10,
    max_features='sqrt',
    max_depth=20,
    random_state=42
)
```

INTERPRETATION AND KEY TAKEAWAYS

Business Implications of this Model:

1. Precisely Adjusted Premiums
2. Avoid Adverse Selection
3. Better Risk Segmentation (leads to Increased Profitability and Portfolio Balance)

- How could we use predicted claim probabilities to segment customers for pricing, retention, or risk management strategies?
- What are the financial consequences of false positives and false negatives in claim prediction?

BUSINESS INSIGHTS

- Prediction Values vs Analyst Interpretation
- Key Challenges Faced Building The Machine Learning Models
- Can This Model Be Used to Predict Lost Costs for Life Insurance Policies?
- Could this model inform reinsurance strategies by identifying high-risk segments?
- How frequently should the model be updated or retrained?

TANK YOU!