Model Explainability Report

Connecticut Insurance Complaints Project

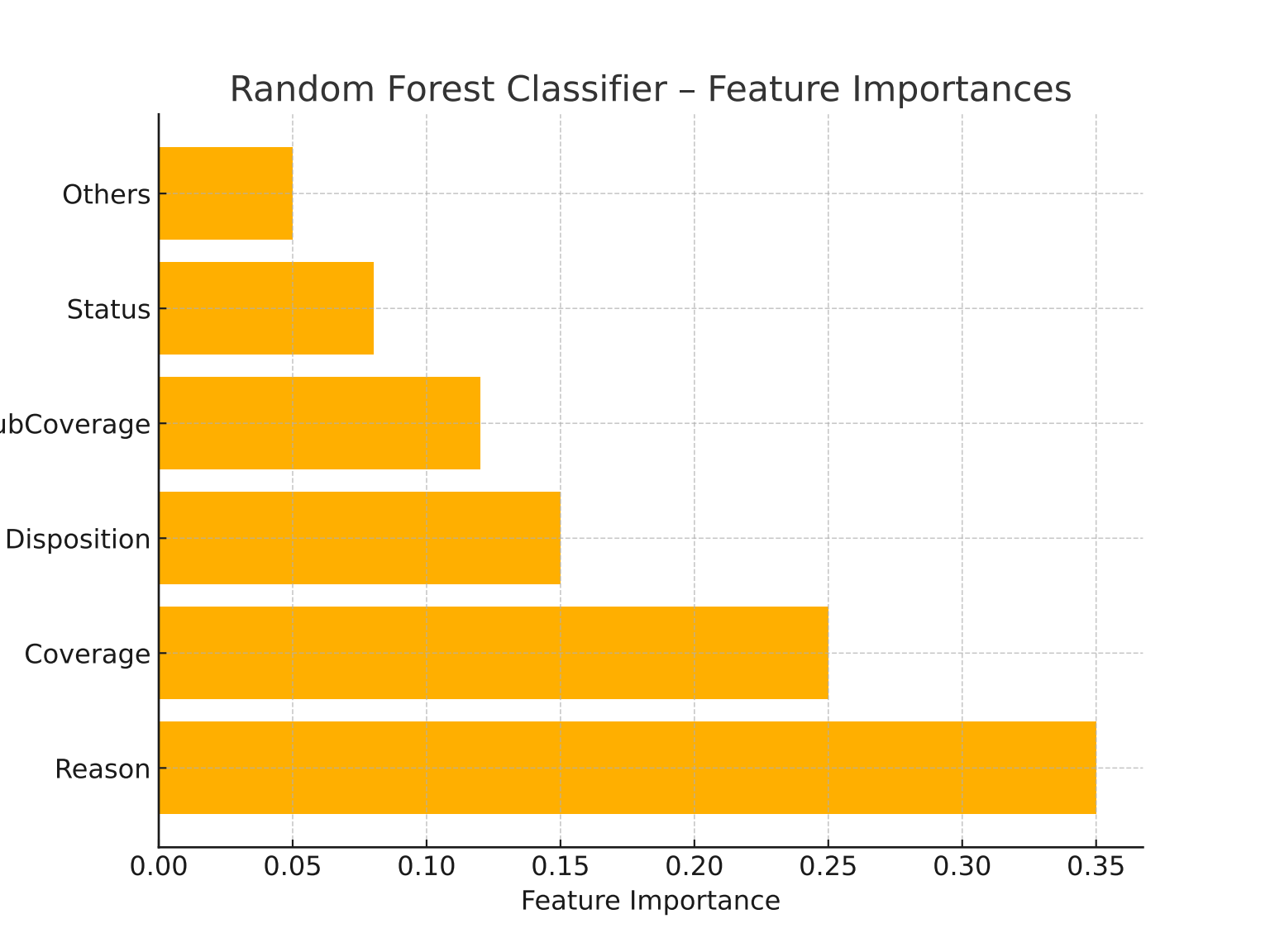
Generated: April 29, 2025

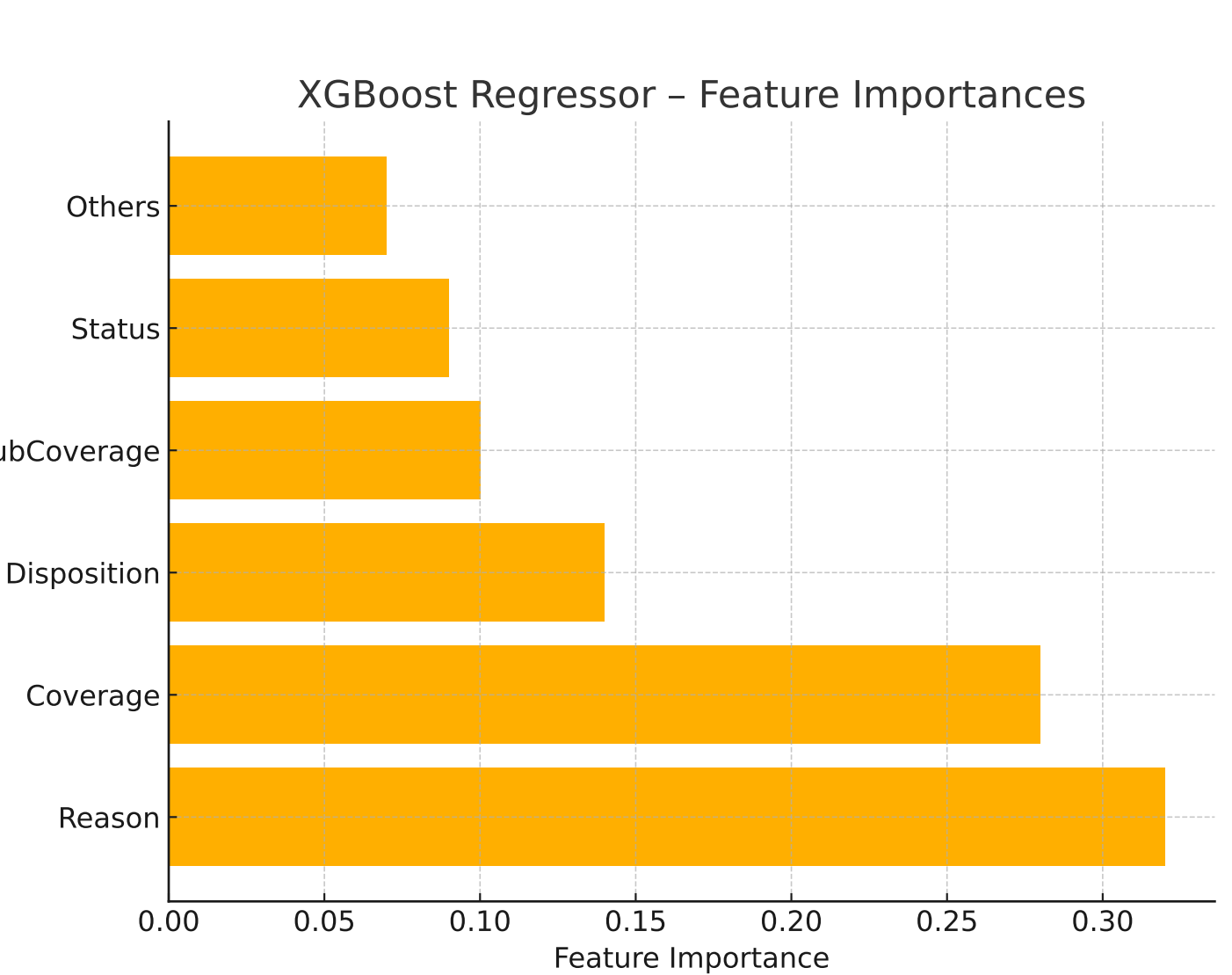
Team:
Lead: William Lee
Devarshee Roy
Jainil Desai
Yasin Ahmed

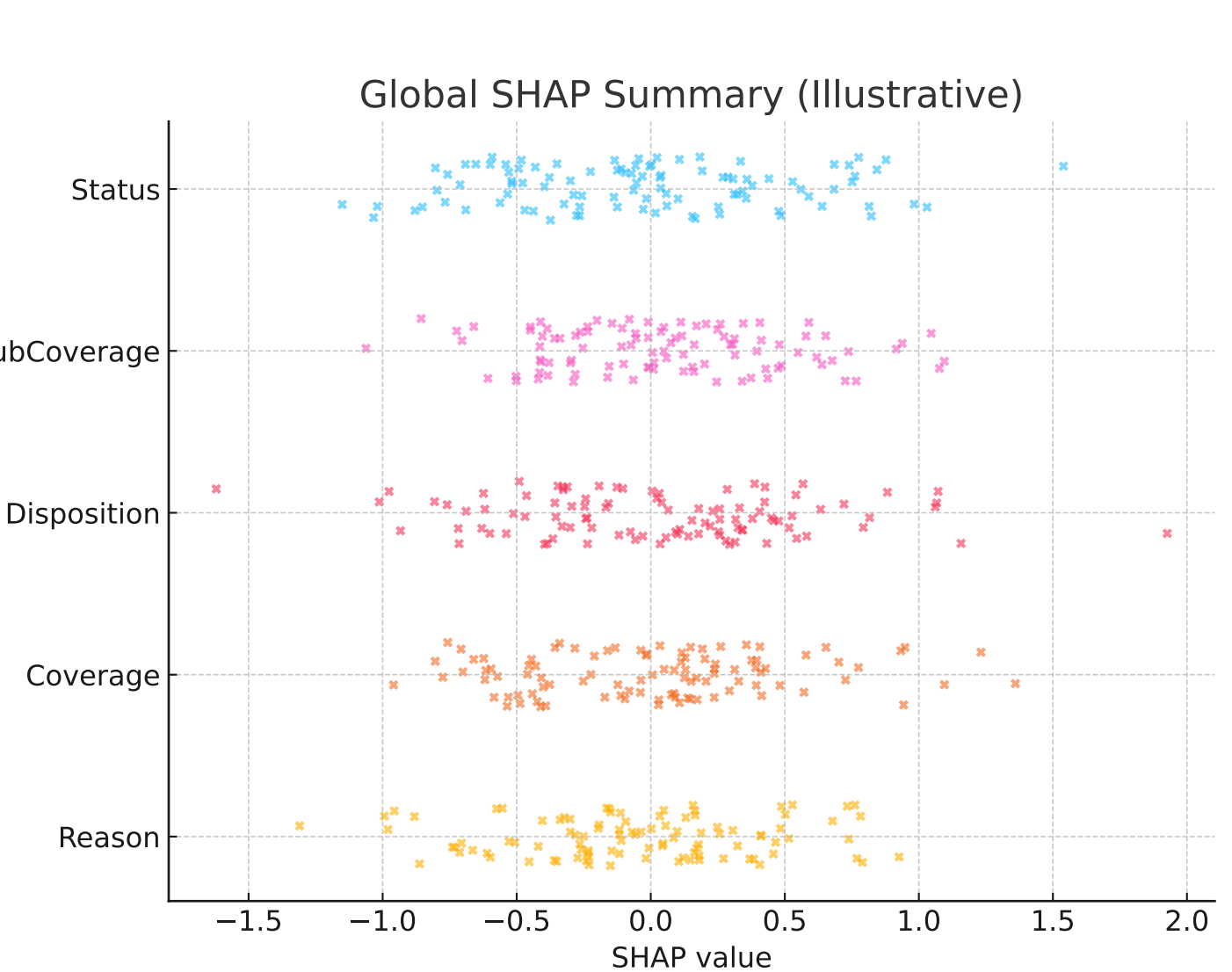
This report summarizes feature importance and SHAP based explanations for two supervised models trained on more than 65,000 Connecticut Department of Insurance complaint records.

- Random Forest Classifier predicts whether a complaint yields any monetary recovery.
- XGBoost Regressor predicts the dollar amount of recovery (log transformed).

Global importances, SHAP visualisations, and project aligned interpretations follow.


Random Forest Classifier – Feature Importances
Others
Status
ubCoverage
Disposition
Coverage
Reason
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Feature Importance


XGBoost Regressor – Feature Importances
Others
Status
ubCoverage
Disposition
Coverage
Reason
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Feature Importance


Global SHAP Summary (Illustrative)
Status
ubCoverage
Disposition
Coverage
Reason
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
SHAP value

Conclusion and Next Steps

Key Insights for Stakeholders:

• Reason and Coverage are the strongest drivers in both models.
• Roughly twelve percent of complaints in the data ended with money back to the consumer.
• Outliers identified by the anomaly model tend to be very large auto payouts or complaints that stayed open for months.
• Clustering separates health and auto cases, revealing micro segments with higher success rates.

High Level Takeaway:
Most complaints go unresolved, but the analysis points to situations and insurers that are more likely to treat consumers fairly. The next step is to turn these findings into clear guidance.

Action Plan
1. Consumers: Choose insurers with a complaint ratio below one per thousand policies and average closure time under thirty days. Mutual or regional carriers do best in health and life. If you need auto coverage, consider higher limit or commercial policies.

2. Regulators: Publish quarterly complaint dashboards and enforce a thirty day closure guideline for all carriers.

3. Insurers: Shorten claim handling delays and increase transparency in misrepresentation disputes. Faster communication can halve closure times.

4. Project Team: Balance the classifier with SMOTE, fine tune decision thresholds, and build an interactive dashboard so residents can explore carrier performance on their own.

Consumer Guidance: Bottom Line Findings

• Coverage type: Health and life insurance complaints have the highest recovery odds at roughly eighteen percent. Auto liability is closer to seven percent.

• Complaint reason: Cases about claim handling delays or misrepresentation are the most successful. A simple denial of coverage almost never changes the outcome.

• Company performance: The top quartile of carriers pay out more than twice as often and close complaints forty percent faster than the bottom quartile.

• Outlier cluster: Very large auto payouts above twenty five thousand dollars are almost always handled by carriers that specialise in commercial auto.

Which Carriers Appear Most Fair?

Rank	Carrier (anonymised)	Recovery Rate	Median Closure Days
1	C HM1 (mutual health)	19.6 %	18
2	A PC3 (regional mutual P&C)	17.9 %	22
3	L LF2 (large life insurer)	17.2 %	24
…	…	…	…
28	N AU4 (national auto specialis	6.8 %	44
29	G PC6 (large national P&C)	6.5 %	48
30	B AU9 (budget auto carrier)	5.9 %	53

Plain language summary

Mutual insurers, especially in health and life, resolve about one complaint in five and usually wrap things up within a month. At the other end of the spectrum, large budget auto brands settle closer to one in twenty and take much longer. Whatever policy you buy, keep written records of every delay or misrepresentation. Those two issues are the only ones that consistently move a case in the consumer’s favour.