Model Development Phase Template

Date	18 June 2024
Team ID	739990
Project Title	Auto Insurance Fraud Detection
Maximum Marks	5 Marks

Feature Selection Report Template

This report provides an overview of the feature selection process for the [Auto Insurance Fraud Detection]. The goal of feature selection is to identify the most relevant and impactful features that contribute to the model's performance.

Feature	Description	Selection (Yes/No)	Reason
months_as_ customer	Number of months the customer has been with the company	Yes	Important for understanding customer loyalty and tenure. |
age	Age of the insured	Yes	Often correlates with risk profiles.
policy_number	Unique identifier for each policy	Yes	Unique Identification number different for each person |
policy_bind_date	Date when the policy was initiated	No	May not directly impact fraud prediction.
policy_state	State where the policy is issued	No	Geographic location can affect claims.
policy_csl	Policy coverage limit |	No	Determines claim amount and severity.
policy_deductable	Deductible amount on the policy	No	Affects claim amounts and patterns.

policy_annual_premium	Annual premium paid	Yes	Directly impacts financial risk assessment.
umbrella_limit	Additional liability coverage limit	No	Indicates extended coverage and risk.
insured_zip	ZIP code of the insured	Yes	Geographic location covered by policy_state.
insured_sex	Gender of the insured	No	Often not a significant factor in claims.
insured_education_level	Education level of the insured	No	May not directly correlate with claims
insured_occupation	Occupation of the insured	No	Occupation can influence risk but may be complex to model.
insured_hobbies	Hobbies of the insured	No	Unlikely to directly impact claims.
insured_relationship	Relationship of the insured with policyholder	No	Typically not a significant factor in claims.
capital-loss	Capital losses recorded	No	Financial data not directly related to insurance claims.
capital-gains	Capital gains recorded	Yes	Financial data directly related to insurance claims.
incident_date	Date of the incident	No	Important for understanding claim timing.
incident_type	Type of incident	No	Affects claim type and severity.
collision_type	Type of collision	No	Affects claim severity and cost.

incident_severity	Severity of the incident	No	Crucial for assessing claim severity.
authorities_contacted	How authorities were contacted	No	Process around claim filing.
incident_state	State where the incident occurred	No	Geographic impact on claims.
incident_city	City where the incident occurred Covered by	No	
incident_hour_of_the_day	Hour of the day when the incident occurred	Yes	Timing can influence claims processing.
number_of_vehicles_involved	Number of vehicles involved in the incident.	Yes	Affects claim severity and complexity
property_damage	Whether property damage was reported	No	Indicates claim type and cost.
bodily_injuries	Number of bodily injuries reported.	Yes	Affects claim severity and cost.
witnesses	Number of witnesses to the incident.	No	Indicates claim complexity and support.
police_report_available	Whether a police report was available.	No	Legal documentation for claims processing.
total_claim_amount	Total amount claimed	Yes	Target variable for prediction. |
injury_claim	Amount claimed for injuries	Yes	Component of total claim amount. |
property_claim	Amount claimed for property damage	No	Component of total claim amount.
vehicle_claim	Amount claimed for vehicle damage	No	Component of total claim amount.
auto_make	Make of the insured	No	Vehicle details impact

	vehicle		claim severity.
auto_model	Model of the insured vehicle	No	Vehicle details impact claim severity.
auto_year	Year of the insured vehicle	Yes	Vehicle age can affect claims.
fraud_reported	Whether fraud was reported.	No	Target variable for model building