INSURANCE CLAIM PREDICTION

▢ Project overview:-

The primary goal of this project is to build the predictive model that help a insurance company which customers are willing to buy a insurance . The dataset contain **595,212 rows** and 59 features and the target column indicate wheathe customer brought the product or not.

▢ Objectives:-

- Task 1:- Build machine learning models to predict which customer will buy the product or not.
- Task 2:- Provide suggestions to the insurance marketing team based on insights derived from the model.

▢ Dataset details:-

- Source:- provided the zip file(PRCP-1010-InsClaimPred.zip)
- Rows:- 595,212.
- Columns:- 59 features .
- Target= Binary value (if purchased=1, not purchased=0)

▢ Data preprocessing :-

- Missing values are handled using median imputation because the dataset is large and anonymus
- Encoding non- numerical fields are factorized.
- Scaled used only Standardscaler only the model that needed it

▢ Model Comparison Summary

Model	Accuracy	F1 Score	Precision	Recall	
Logistic Regression	0.963165	0.00000	0.00000	0.0000	
Decision Tree	0.918030	0.047194	0.063854	0.054274	
Random Forest	**0.963156**	**0.00000**	0.00000	0.00000	
Gradient Boosting	0.963148	0.333333	0.000456	0.000911	

- Random Forest gave the best compare to remaing models .

▢ Business Insights & Marketing Suggestions:-

Business strategic recommendations:

- Focus on high-probability customers:-

To increase the customers uses the models to generate the customer probability and focus on the customers with high predicted score.

- Improve follow-up strategy for borderline customers :-

To improve strategy for customers give them the discounts for them and give the offer and cashback and also send them the personlized reminders for the customers.

- Personalize communication:-

Use customer segmentation based on model output as for younger customer use the digital ads and for High risk customers use the insurance awareness campaigns and finally for existing policyholderes use the cross selling strategy.

- Optimize marketing channels :-

For marketing channels we can use the emails, sms, mobile apps and lastly call center follow up.

▢ Final Model Recommendation :-

The recommended model for deployment is ‘’Random Forest ‘’

Reasons:-

1. It has the highest value in all compare to remaing models .
2. This model has fastest traning of the dataset.
3. This also handle the large dataset

▢ Conclusion :-

This project successfully built a robust prediction system for customer purchase behavior in insurance. Despite the lack of feature names, the model provides strong predictive power and actionable insights.

The final output includes:

- Trained ML models
- Model comparison metrics
- Feature importance list
- Business recommendations
- Challenges & solutions

The insurance company can now integrate this model into their CRM system to improve sales targeting and boost conversion rates.