AI-Powered Intelligent Insurance Risk Assessment and Customer Insights System

**Executive Summary:**

Objective:

To build an AI-powered system that enhances insurance risk classification, detects fraudulent, claims, segments customers, and extracts insights from feedback using NLP and ML techniques.

Solution Overview:

- Risk Classification: Classify policyholders into low/medium/high risk.

- Fraud Detection: Flag potentially fraudulent insurance claims.

- Customer Segmentation: Cluster customers based on behavior & demographics.

- Sentiment Analysis: Understand customer feedback using NLP.

- NLP Tasks: Multilingual summarization & translation of policy documents.

**Exploratory Data Analysis (EDA)**

Key Visualizations & Insights:

- Claim amount distribution -> Detect outliers.

- Correlation heatmaps -> Identify key influencing variables.

- Customer demographics by region -> Insights for segmentation.

- Sentiment trends from feedback over time.

- Fraudulent vs non-fraudulent feature patterns.

**Tools used:**

- Python (Pandas, Matplotlib, Seaborn, Plotly)

- NLP: WordClouds, N-grams, Sentiment Scores

**Model Training & Evaluation**

Modules and Algorithms Used:

Task | Models | Metrics

-------------------------|----------------------------------------------------------------|--------------------------

Risk Classification | Random Forest, XGBoost, Logistic Regression | Accuracy, AUC-ROC

Fraud Detection | XG Boost, Isolation Forest, SVM | Precision, Recall, F1-score

Sentiment Analysis | TF-IDF + Logistic Regression / BERT | Accuracy, F1-score

Customer Segmentation | K Means, DBSCAN | Silhouette Score

NLP Tasks | Hugging Face Transformers (BART, m BART) | BLEU, ROUGE

Model Comparison Results:

- XG Boost achieved highest AUC-ROC for fraud detection.

- K Means with 4 clusters yielded meaningful customer segments.

- BERT-based model outperformed traditional models for sentiment.

**Challenges Faced & Improvements**

Challenge | Solution

------------------------------- | ----------------------------------------

Imbalanced fraud data | Used SMOTE & class weights

Multilingual documents | Leveraged m BART & Marian MT for translation

Noisy customer feedback | Applied text cleaning, stop word removal, lemmatization

Model interpretability | Added SHAP for explaining predictions

Data Privacy | Masked personal identifiable information (PII)

**Future Enhancements**

- Deploy model APIs with FastAPI for modular backend.

- Add real-time fraud alert system via Kafka + Flask.

- Integrate with insurance CRMs for continuous feedback ingestion.

- Improve multilingual NLP coverage (add more Indic/Asian languages).

- Include geospatial analysis of fraud/risk based on location.

- Use AutoML for ongoing hyperparameter tuning.

**Deployment Instructions**

Environment Setup:

$ git clone https://github.com/your-username/insurance-ai-system.git

$ cd insurance-ai-system

$ python -m venv venv

$ source venv/bin/activate

$ pip install -r requirements.txt

Running the Streamlit App:

$ cd deployment

$ streamlit run app.py

Backend API:

$ uvicorn api:app --reload

Docker Deployment:

FROM python:3.10

COPY . /app

WORKDIR /app

RUN pip install -r requirements.txt

CMD ["streamlit", "run", "deployment/app.py"

THANK YOU