Detecció de Fraus en Targetes de Crèdit Utilitzant Tècniques d'Aprenentatge Automàtic

Marc Serra Ortega

Abstract—La detecció de fraus en transaccions amb targetes de crèdit és essencial per a la seguretat financera. Aquest estudi aborda aquesta problemàtica mitjançant l'ús de tècniques d'aprenentatge autoaticen un conjunt de dades altament desequilibrat. Amb 0.172% de fraus en 284,807 transaccions, la detecció eficaç és un desafiament. Es presenta un enfocament utilitzant models d'aprenentatge automatic combinats amb l'utilitzacio de técnicas diverses per tractar el desbalanceig. El conjunt de dades, proporcionat per Worldline i la Université Libre de Bruxelles, s'analitza amb mètriques com la puntuació F1 macro i la corba Precision-Recall. El treball contribueix a la recerca en detecció de fraus i destaca la necessitat de l'evaluacio de de manera correcta i robusta en situacions desequilibrades.

Keywords—Credit Card Fraud Detection, Unbalanced Classification, Oversampeling, Undersapling

1 Introducció

Amb l'augment exponencial de transaccions amb targetes de crèdit, la detecció eficient de frau esdevé essencial per a la integritat del sistema financer. El desequilibri inherent en aquestes dades, amb un nombre significativament més alt de transaccions no fraudulentes que fraudulentes, planteja reptes substancials en la creació de models de detecció precisos. Aquest document se centra en l'ús de tècniques innovadores, com l'Aprenentatge Sensible al Cost i el remostreig, per abordar aquesta desproporció. A través de l'anàlisi detallada de diferents estratègies, s'identifica un model òptim, un Voting Classifier amb aprenentatge sensible al cost, que mostra un rendiment superior en la detecció de frau. Aquesta investigació no només destaca l'eficàcia d'aquest model, sinó que també posa de manifest les limitacions existents i ofereix perspectives per a futures millores en la detecció de frau en el context específic de les transaccions amb targetes de crèdit.

2 Proposta/metodologia

Per endinsar-nos a les complexitats del conjunt de dades, és crucial entendre que les característiques numèriques provenen d'una transformació PCA, que inclou els components principals V1 a V28. No obstant això, és destacable que les variables "Temps" i "Import" no han estat sotmeses a aquesta transformació específica.

Amb inici a l'anàlisi exploratori de dades, establim tres objectius principals:

1. Explorar la Relació Entre el Temps i les Transaccions Fraudulentes: Volem investigar possibles patrons en l'aspecte temporal de les transaccions fraudulentes. Això implica una anàlisi exhaustiva de les dades i una exploració de possibles correlacions amb el temps.
2. Analitzar la Relació Entre l'Import de les Transaccions i el Fraude: La nostra atenció es centra en comprendre si les transaccions fraudulentes tendeixen a mostrar imports més grans o més petits en comparació amb les legítimes. Destacablement, les disparitats significatives en aquest aspecte indiquen una prevalença d'imports més elevats associats amb les transaccions fraudulentes.
3. Identificar i Analitzar Outliers al Conjunt de Dades: Com que els outliers poden afectar significativament la integritat de la nostra anàlisi, identificarem i analitzarem les seves característiques. És important destacar que no hem optat per eliminar aquests outliers a causa de la seva estreta associació amb instàncies de frau.

El desequilibri inherent en el conjunt de dades requereix l'aplicació de diverses tècniques per abordar aquest probleme, i en aquesta secció explorarem els següents mètodes:

- **Undersampling**: Utilitzant tècniques com RandomUnderSampler i ClusterCentroids.
- **Oversampling:** Emprant metodologies com SMOTE i ADASYN.
- **Combinació d'Oversampling i Undersampling**: Investigant l'eficàcia de SMOTEENN en la compensacio de Overfitting i Underfiting.
- **Aprenentatge Sensible al Cost:** Endinsant-nos en models sensibles al cost associat amb la mala classificació.

Un cop implementades aquestes tècniques, procedirem amb una anàlisi detinguda dels resultats, avaluant la seva eficàcia en la detecció de frau en targetes de crèdit mitjançant la validació creuada. Donada la naturalesa desbalancejada del nostre conjunt de dades i l'objectiu primordial de la detecció de frau, concedim especial importància a la mètrica F1 Macro i a les corbes PR. És fonamental abordar aquesta desproporció amb cautela per evitar la creació de models amb sobreajustament, una preocupació que guia les estratègies de remostreig implementades.

A continuació, dirigirem la nostra atenció a l'avaluació dels diferents models utilitzats en la detecció de frau en targetes de crèdit. Les principals mètriques de rendiment, com la precisió, la sensibilitat, la puntuació F1 i l'àrea sota la corba de precisió-recall (PRAUC), seran escrutinades. L'anàlisi comparativa dels resultats obtinguts amb cada model ens guiarà en la selecció del model més eficaç per a la detecció de frau en targetes de crèdit.

Concloent la nostra anàlisi, realitzarem una anàlisi a fons del model final i extreurem conclusions rellevants. L'avaluació del rendiment del model en el conjunt de proves serà exhaustiva, i es discutiran les limitacions, així com es proporcionaran suggeriments per a possibles millores. En resum, aquesta anàlisi científica sobre la detecció de frau en targetes de crèdit culminarà en una visió coherent i completa dels resultats i les implicacions.

3 Experiments, Resultats i Anàlisi

La metodologia triada implica la experimentació amb diferents estratègies d'aprenentatge desbalancejat. La utilització de tècniques de mostreig es van mostrar prometedores en un principi per la possibilitat que presenten en generar un conjunt de dades representatiu i equilibrat. Les primeres proves amb Under-

Sampling van començar mostran conjunts de dades amb una gran separabilitat tant per el Random Under Sampler, un mètode de *Prototype Selection*, tant com per Cluster Centroids que utilitza Prototype Generation. Tot i que finalment van resultar en un en mal rendiment en la validació creuada degut a la pèrdua d'informació resultant de la gran decaiguda en mostres per a la classe majoritària, obtenint una F1 Macro de 0.537 per a KNN amb Random Under Sampler i de 0.614 per a KNN amb Cluster Centroids.

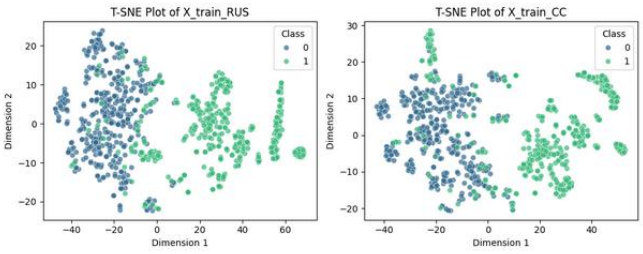

Fig. 1. Gràfic de la representació T-SNE per als conjunts de dades generats per els models d'Under Sampling Random Under Sampler i Cluster Centroids

A continuació es va experimentar amb les estratègies d'Over Sampling per tal de generar mostres per a la classe minoritaria de forma representativa. Per aquest objectiu es van triar els algoritmes SMOTE, que genera mostres sintètiques en la regió entre punts de la classe minoritària, i ADASYN una variant del SMOTE que adapta el nivell d'Over Sampling per a cada mostra, basant-se en la seva dificultat de classificació. Aquests models si van presentar un bon rendiment general obtenint una F1 Macro de 0.924 per al Random Forest amb SMOTE i 0.922 per al Random Forest amb ADASYN. Tambes es van utilitzar corbes PR per entendre millor el compromís entre precisió i recall el qual va mostrar l'abrupte baixada en precisió a partir d'un valor de recall de aproximadament 0.82, expresant la posible limitació que presentaven els models per a classificar certs exemples fraudulents.

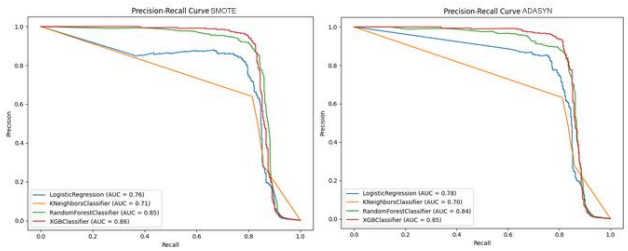

Fig. 2. Grafica PR per els diferents models avaluats amb SMOTE (gráfica de la dreta) i ADASYN (gráfica de la esquerra).

Seguidament s'analitza la combinació d'Over Sampling i Under Sampling per tal de mitigar el possible soroll generat per les estratègies d'Over Sampling. Per aquesta tasca ens va triar l'algoritme SMOTEENN una combinació del mètode SMOTE i ENN (Edited Nearest Neighbors) un algoritme d'Under Sampling. Amb aquest enfocament es no va millorar el rendiment obtingut per SMOTE indicant que possiblement ENN eliminava exemples clau del tipus *outlaier* corresponents a exemples de frau clau per a classificar algunes de les mostres del conjunt de dades, tot i que si que es va aconseguir un bon rendiment en general. Amb aquesta estrategia es va aconseguir una F1 Macro de 0.91 per al Random Forest.

Finalment es va experimentar amb Aprenentatge Sensible al Cost per tal de modificar el pes de les funcions de cost per els diferents models triats a entrenar i així abordar el gran desbalanceig en el conjunt de dades. El resultat d'aquest enfocament va resultar molt satisfactori, donant un molt bon rendiment en tots els models pero , especialment en el Voting Classifier que va derivar de la combinació dels anteriors models per tal d'obtenir un classificador amb una major capacitat de generalitzar. Aquest model ha obtingut el resultat més alt en la validació creuada amb una F1 Macro de 0.938 i una exelent rendiment en la majoria de models menys en la Logistic Regression.

Taula F1 Models Simples

	f1_macro
LogisticRegression_-8303857189465698474	0.561600
KNeighborsClassifier_-3336143508539130970	0.909363
RandomForestClassifier_1858126956338585078	0.919852
XGBClassifier_-2781219804619487201	0.931367

Taula F1 Voting Classifier

	f1_macro
VotingClassifier	0.938196

Fig. 3. Taules F1 Macro per els diferents models avaluats amb Aprenentatge Sensible al Cost.

La utilització del Voting Classifier dotat d'aprenentatge sensible al cost,aconsegueix un gran rendiment destacant l'alta F1 Macro de 0.925 en el conjunt de testeig confirmant així l'elecció coherent amb la validació creuada com a classificador per a detectar el frau en el conjunt de dades donat. A mes a mes, l'accuracy pràcticament perfecte de 0.999 i la precisió de 0.927 recalquen la aquest fet. Per altre banda, tot i que la Recall no es baixa, es la métrica a millorar amb un valor de 0.785.

	Precision	Recall	F1-Score	Support
Class 0	0.9996	0.9999	0.9998	56864
Class 1	0.9277	0.7857	0.8508	98
Accuracy			0.9995	56962
Macro Avg	0.9637	0.8928	0.9253	56962
Weighted Avg	0.9995	0.9995	0.9995	56962

Taula 1. Informe de classificació del model final

La utilitzacio les corbes PR per determinar el bon rendiment dels models donada la naturaleza desbalanceada del conjunt de dades ha sigut clau per analitzar el rendiment general del model, donant una PRAUC de 0.86. A mes a mes aquesta grafica ha estat clau per analitzare compromís entre la precisió i la recall en els diferents models amb que s'a experimentat. Destacant una abrupte caiguda de la precisió al voltant d'un valor de Recall de 0.83. Aquest comportament ens indica que hi ha una petita quantitat d'exemples que son molt difícils de classificar per la semblança de característiques que comparteixen amb els exemples no fraudulents.

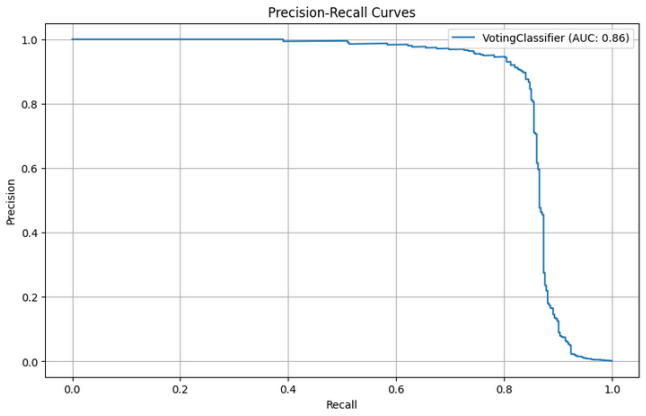

Fig. 4. Gràfic de la corba PR (Precision-Recall) del model Voting Classifier.

Analitzant les caracteristiques amb mes importancia per al Voting Classifier observem que la característica més rellevant per al nostre model és la V14, amb una diferència significativa. A més, en analitzar les dues característiques més importants (V14 i V10) mitjançant un joinplot, es destaca una notable separabilitat entre les classes. Aquesta observació suggereix que l'ús combinat d'aquestes dues característiques derivades del PCA mostra una distinció clara entre la majoria de transaccions legítimes i fraudulentes. Aquesta separabilitat pot indicar que les característiques V14 i V10 aporten una quantitat substancial d'informació al model per a la detecció de frau en les transaccions bancàries.

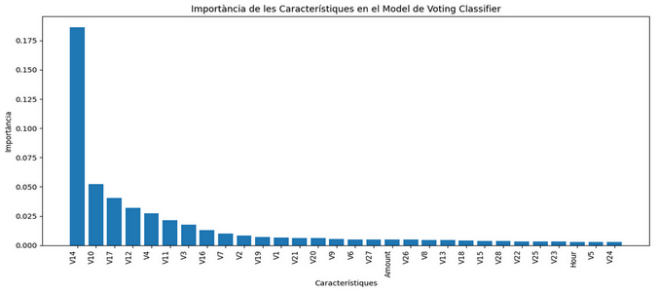

Fig. 5. Histogramade Importancia de caracteristiques model Voting Classifier.

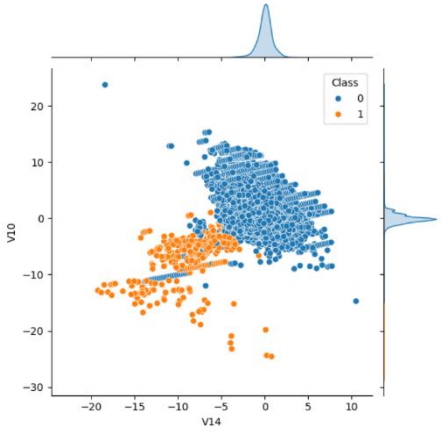

Fig. 6. Join plot de les caracteristiques V14 i V10.

Com a últim phay que destacarnalitzar el Voting Classifier, ens fixem en la matriu de confusió i observem que la majoria d'errors, com s'havia prèviament predit, ocórren en exemples fraudulents (Class 1) que es classifiquen de manera erronia. Aquest fenomen podria ser resultat de dues raons principals i ens fa preguntarnos quina es la naturaleza i quin es l'efecte d'aquest exemples de frau mal classificats en la banca.

- Potser aquests exemples van ser etiquetats incorrectament des del principi.
- Potser els estafadors utilitzen tècniques molt avançades per cometre frau, de manera que les transaccions fraudulentas siguin pràcticament indistingibles de les transaccions legítimes.

A continuació, investigarem les transaccions mal classificades i l'impacte que aquestes poden tenir per a la banca.

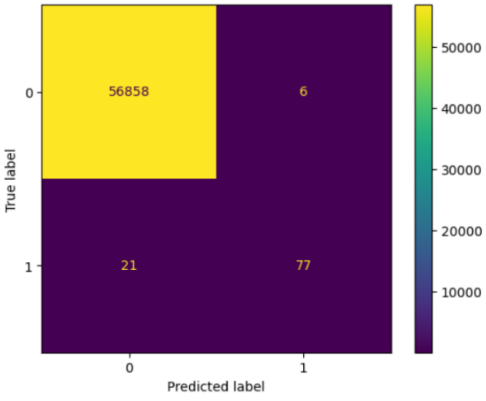

Fig. 7. Matriu de confusió model Voting Classifier.

Observant els exemples de frau mal classificats descobrim que hi ha diverses transaccions amb un valor elevat (superior a 500) , i si aquestes transaccions passessin desapercebudes, podrien ocasionar un cost considerable per a la banca. Seguidament vamfixarnos en la mitjana de les transaccions fraudulentes que el model no ha pogut classificar correctament mostrant valors molt més semblants a les transaccions legítimes, especialment en les característiques amb més pes per a detectar frau, com ara V14, V10, V17 i V12, que ara s'apropen al valor mitjà de les transaccions no fraudulentes. A més a més, observem que aquest efectivamente, de mitjana, aquest tipus de transaccions són particularment costoses per al banc, superant el valor mitja de les transaccions fraudulentes "simples" i legitimes per una quantitat significativa.

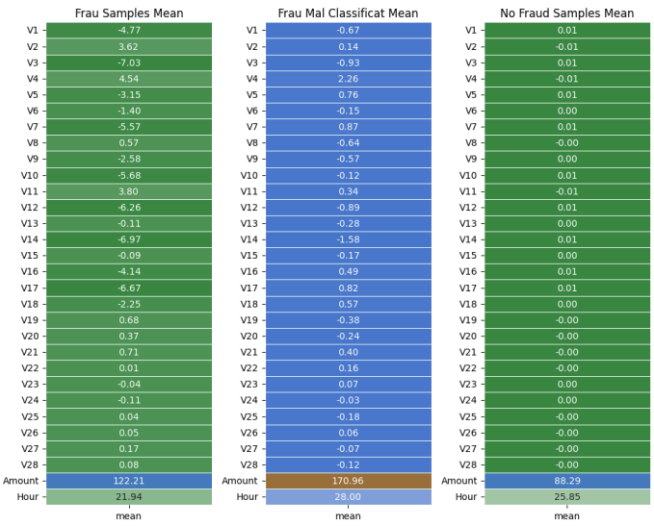

Fig. 8. Comparacio de carateristiques mitjanes per a (d'esquerra a dreta) exeples de frau "simples", exemples de frau mal classificats i exemples legitims.

5 Conclusions

En aquesta anàlisi detallada, hem explorat amb profunditat les tècniques i estratègies implementades per fer front al desafiament de la detecció de frau en un conjunt de dades de transaccions amb targeta de crèdit, caracteritzat per un desequilibri significatiu.

Pel que fa al rendiment del model, hem optat per un enfocament centrat en un Voting Classifier amb aprenentatge sensible al cost, que integra models fonamentals com Random Forest, Logistic Regression, KNN i XGBoost. El resultat final, mesurat amb la mètrica F1 Macro, destaca amb un valor de 0.925.

És notable la capacitat del model per classificar la majoria dels exemples amb una alta precisió, amb una precisió del 99,9% per a la classe legítima (Class 0) i del 92,7% per a les instàncies fraudulentes (Class 1). Tot i així, cal assenyalar la limitació del model en superar una recall superior al 82% sense comprometre el rendiment global. Aquest comportament pot ser atribuït a la complexitat inherent d'algunes transaccions o a tècniques sofisticades utilitzades pels estafadors, dificultant la distinció entre un frau i una transacció vàlida.

Les implicacions pràctiques d'aquest model eficient són significatives per a la prevenció de pèrdues econòmiques associades a transaccions fraudulentes. La identificació precisa de transaccions sospitoses és crucial per a les institucions financeres, contribuint a gestionar pèrdues i preservar la confiança dels clients.

Pel que fa als suggeriments per a futures investigacions i millores, es recomana explorar la combinació de models especialitzats en diferents tipus de frau i tècniques d'aprenentatge no supervisat, com ara autoencoders, per abordar transaccions fraudulentes difícils de distinguir. Millorar la interpretabilitat del model amb dades no encriptades és una àrea crítica per a una comprensió més profunda del seu comportament.

En conclusió, la combinació d'un Voting Classifier amb aprenentatge sensible al cost ha demostrat ser una solució robusta per a la detecció de frau en transaccions amb targeta de crèdit. Malgrat les limitacions identificades, aquest enfocament ha superat altres estratègies, oferint consideracions pràctiques significatives per a les institucions financeres. El futur de la recerca implica l'exploració de models més especialitzats i millores en la interpretabilitat per abordar les limitacions actuals del sistema

Bibliografia (Opcional)

[1] O. Russakovskyet al., "ImageNet large scale visual recognition challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classificationwith deep convolutional neural networks," inProc. Int. Conf. Neural Inf.Process. Syst., 2012, pp. 1097–1105.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks forlarge-scale image recognition," inProc. Int. Conf. Learn. Representations,2014.

[4] C. Szegedyet al., "Going deeper with convolutions," inProc. IEEE Conf.Comput. Vis. Pattern Recognit., Jun. 2015, pp. 1–9.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for imagerecognition," inProc. IEEEConf. Comput. Vis. PatternRecognit., 2016, pp. 770–778.