



UNIVERSIDADE PRESBITERIANA MACKENZIE

PROJETO APLICADO I

DETECÇÃO DE FRAUDES EM TRANSAÇÕES DE CARTÃO DE  
CRÉDITO

Déborah Silvério Alves Morales – RA 10728563

Diógenes Nimário de Araújo Pereira – RA 10424898

Lucas Iglezias dos Anjos – RA 10433522

Luiz Benlardi Neto – RA 10724617

São Paulo

2025

## Sumário

Resumo.....	4
Lista de Figuras.....	5
<b>Figura 1 - Distribuição das Classes no Dataset:</b> Este gráfico de barras mostra a quantidade de transações legítimas (Class = 0) e fraudulentas (Class = 1) no dataset. É evidente o forte desbalanceamento: as transações legítimas representam a imensa maioria, enquanto as fraudulentas são uma fração mínima. Essa característica exige técnicas específicas de modelagem para evitar viés nos resultados.....	5
<b>Figura 2 - Distribuição dos Valores por Classe:</b> O boxplot compara os valores (Amount) das transações legítimas e fraudulentas. É possível observar que transações fraudulentas tendem a ter valores mais concentrados, enquanto as legítimas apresentam maior variabilidade. Essa diferença pode ser explorada como um possível padrão para detecção.....	6
<b>Figura 3 - Mapa de Correlação entre Variáveis:</b> Este heatmap exhibe a correlação entre todas as variáveis do dataset. As variáveis V1 a V28 foram transformadas por PCA, então não possuem significado direto, mas o mapa ajuda a identificar relações fortes ou fracas entre elas. Isso é útil para entender a estrutura dos dados e selecionar variáveis relevantes para a modelagem.....	7
<b>Figura 4 - Distribuição Temporal das Transações por Classe:</b> O histograma mostra como as transações estão distribuídas ao longo do tempo (Time), separadas por classe. Embora a maioria das transações seja legítima, é possível observar em quais períodos as fraudes ocorrem com maior frequência, o que pode indicar padrões temporais relevantes.....	7
<b>Figura 5 - Distribuição dos Valores das Transações:</b> Este histograma mostra a frequência dos valores (Amount) das transações. A maioria das transações possui valores baixos, com poucos casos de valores altos. Essa distribuição ajuda a entender o perfil financeiro das transações e pode ser usada para identificar outliers.....	8
<b>Figura 6 - Visualização PCA das Transações:</b> Este gráfico de dispersão mostra a projeção das variáveis V1 a V28 em duas dimensões principais (PC1 e PC2), usando PCA. As cores representam as classes. Embora haja sobreposição, é possível notar agrupamentos distintos, sugerindo que o modelo pode aprender a separar as classes com base nessas variáveis.....	8
.....	9
<b>Figura 7 - Volume Acumulado de Transações ao Longo do Tempo:</b> Este gráfico de linha mostra o crescimento acumulado do número de transações conforme o tempo avança. Ele ajuda a visualizar o ritmo de atividade no dataset e pode revelar picos ou padrões de uso que influenciam a ocorrência de fraudes. ....	9
Lista de Tabelas .....	10
<b>Tabela 1 - Amostra das Transações do Dataset</b> .....	10
<b>Tabela 2 - Estatísticas Descritivas por Classe</b> .....	10
<b>Tabela 3 - Frequência de Transações por Classe</b> .....	10
Introdução.....	11
Definição do Problema.....	12
Contextualização da Empresa.....	13
Descrição do Dataset.....	14
Metodologia Proposta.....	15



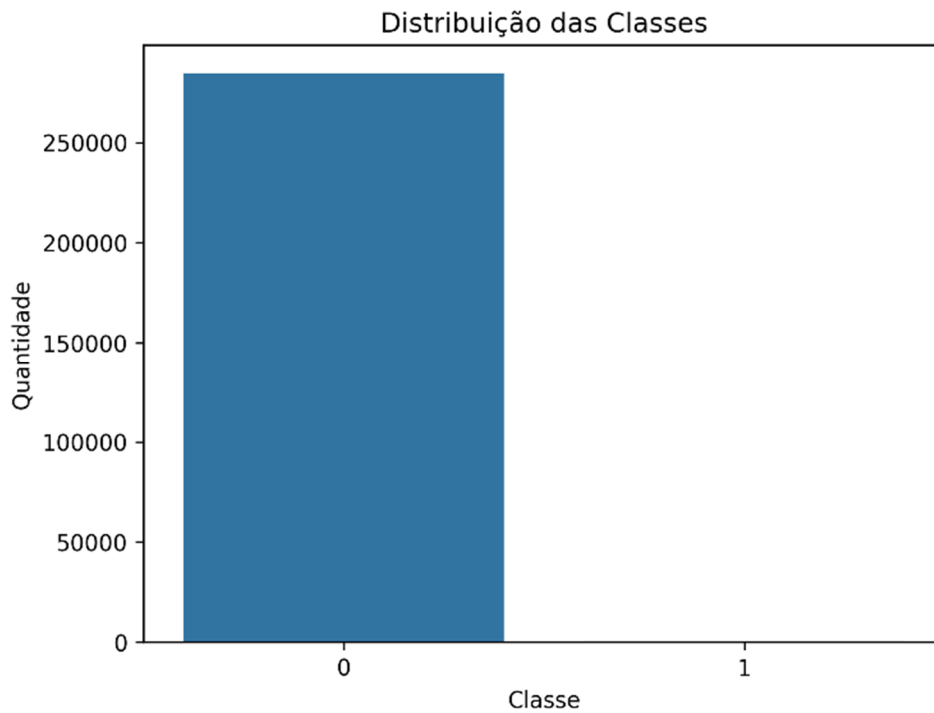
<b>Cronograma.....</b>	<b>17</b>
<b>Repositório do Projeto.....</b>	<b>17</b>
<b>Referências.....</b>	<b>17</b>

## Resumo

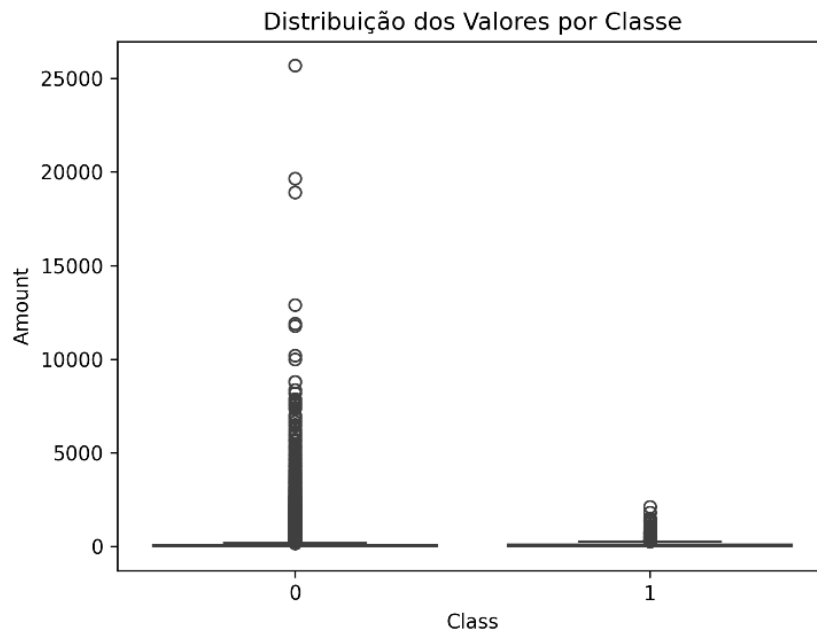
Este projeto tem como objetivo desenvolver um modelo preditivo para detecção de fraudes em transações de cartão de crédito, utilizando técnicas de ciência de dados e aprendizado de máquina. O conjunto de dados utilizado é o Credit Card Fraud Detection, que apresenta forte desbalanceamento entre transações legítimas e fraudulentas. Foram aplicadas etapas de análise exploratória, pré-processamento, modelagem e avaliação utilizando diversos algoritmos. Os resultados esperados incluem a identificação eficaz de fraudes, contribuindo para a segurança das operações financeiras.

**Palavras-chave:** detecção de fraudes, aprendizado de máquina, análise de dados, cartão de crédito, desbalanceamento.

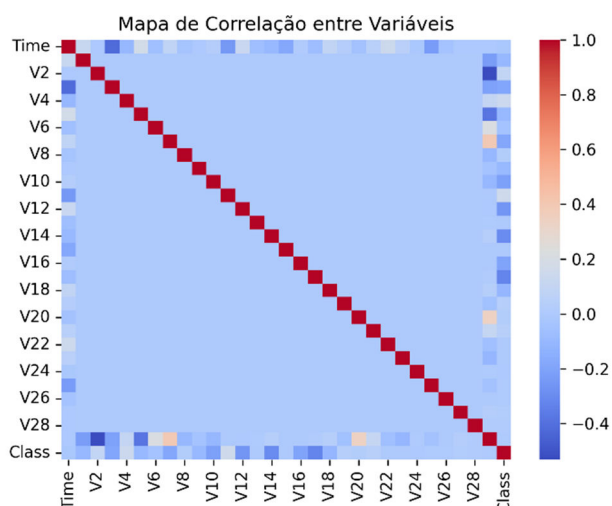
## Lista de Figuras



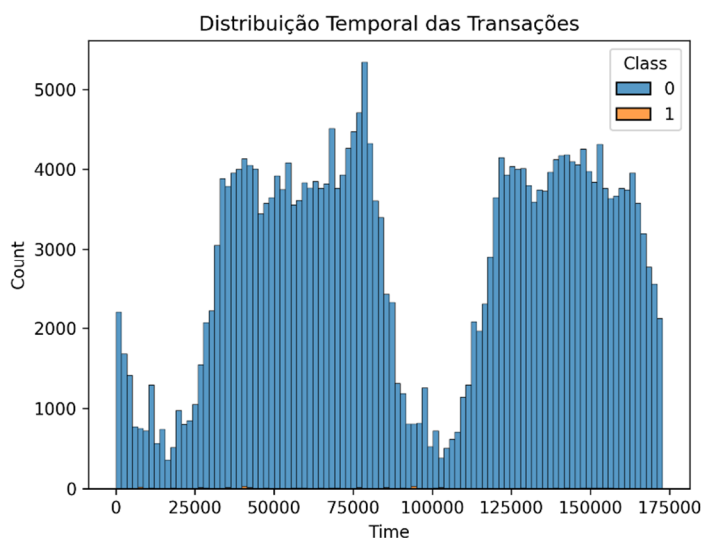
**Figura 1 - Distribuição das Classes no Dataset:** Este gráfico de barras mostra a quantidade de transações legítimas (Class = 0) e fraudulentas (Class = 1) no dataset. É evidente o forte desbalanceamento: as transações legítimas representam a imensa maioria, enquanto as fraudulentas são uma fração mínima. Essa característica exige técnicas específicas de modelagem para evitar viés nos resultados.



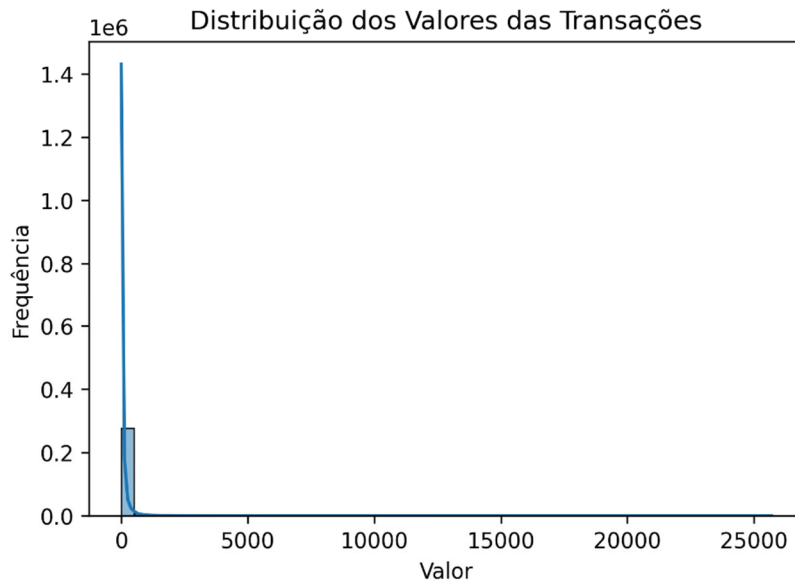
**Figura 2 - Distribuição dos Valores por Classe:** O boxplot compara os valores (Amount) das transações legítimas e fraudulentas. É possível observar que transações fraudulentas tendem a ter valores mais concentrados, enquanto as legítimas apresentam maior variabilidade. Essa diferença pode ser explorada como um possível padrão para detecção.



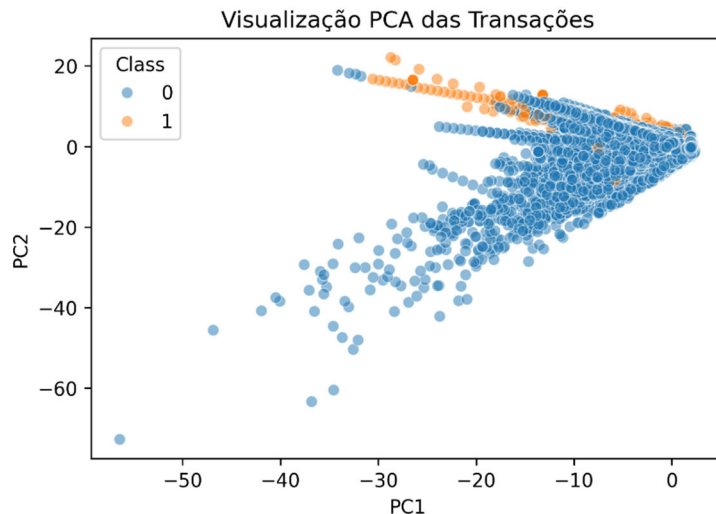
**Figura 3 - Mapa de Correlação entre Variáveis:** Este heatmap exibe a correlação entre todas as variáveis do dataset. As variáveis V1 a V28 foram transformadas por PCA, então não possuem significado direto, mas o mapa ajuda a identificar relações fortes ou fracas entre elas. Isso é útil para entender a estrutura dos dados e selecionar variáveis relevantes para a modelagem.



**Figura 4 - Distribuição Temporal das Transações por Classe:** O histograma mostra como as transações estão distribuídas ao longo do tempo (Time), separadas por classe. Embora a maioria das transações seja legítima, é possível observar em quais períodos as fraudes ocorrem com maior frequência, o que pode indicar padrões temporais relevantes.

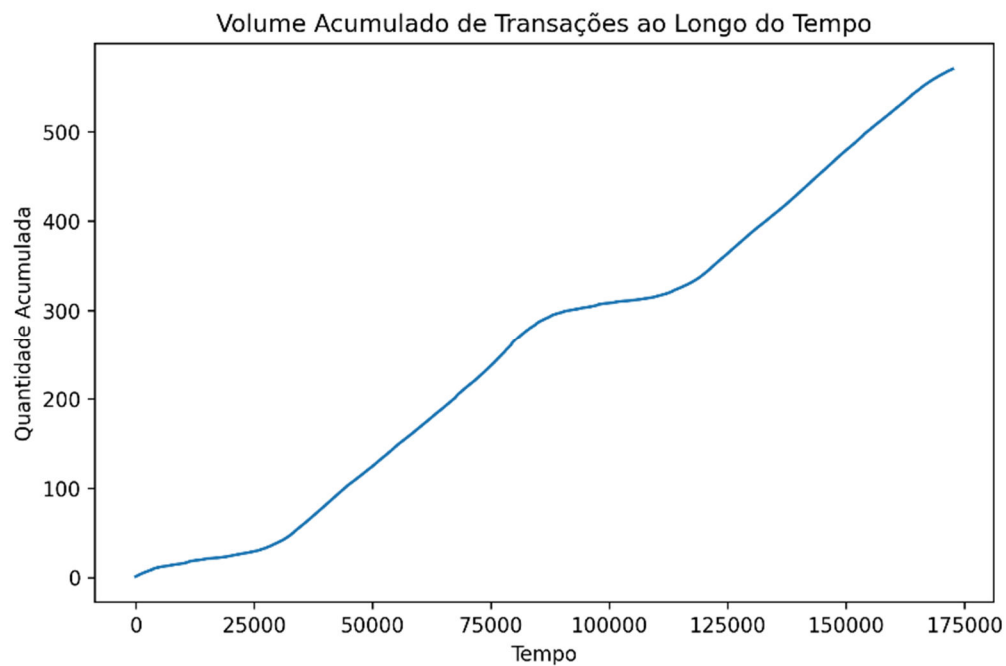


**Figura 5 - Distribuição dos Valores das Transações:** Este histograma mostra a frequência dos valores (Amount) das transações. A maioria das transações possui valores baixos, com poucos casos de valores altos. Essa distribuição ajuda a entender o perfil financeiro das transações e pode ser usada para identificar outliers.



**Figura 6 - Visualização PCA das Transações:** Este gráfico de dispersão mostra a projeção das variáveis V1 a V28 em duas dimensões principais (PC1 e PC2), usando PCA. As cores representam as classes. Embora haja sobreposição, é possível notar agrupamentos distintos, sugerindo que o modelo pode aprender a separar as classes com base nessas variáveis.





**Figura 7 - Volume Acumulado de Transações ao Longo do Tempo:** Este gráfico de linha mostra o crescimento acumulado do número de transações conforme o tempo avança. Ele ajuda a visualizar o ritmo de atividade no dataset e pode revelar picos ou padrões de uso que influenciam a ocorrência de fraudes.

## Lista de Tabelas

**Tabela 1 - Amostra das Transações do Dataset**

Time	Amount	Class	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36	V37	V38	V39	V40	V41	V42	V43	V44	V45	V46	V47	V48	V49	V50	V51	V52	V53	V54	V55	V56	V57	V58	V59	V60	V61	V62	V63	V64	V65	V66	V67	V68	V69	V70	V71	V72	V73	V74	V75	V76	V77	V78	V79	V80	V81	V82	V83	V84	V85	V86	V87	V88	V89	V90	V91	V92	V93	V94	V95	V96	V97	V98	V99	V100	V101	V102	V103	V104	V105	V106	V107	V108	V109	V110	V111	V112	V113	V114	V115	V116	V117	V118	V119	V120	V121	V122	V123	V124	V125	V126	V127	V128	V129	V130	V131	V132	V133	V134	V135	V136	V137	V138	V139	V140	V141	V142	V143	V144	V145	V146	V147	V148	V149	V150	V151	V152	V153	V154	V155	V156	V157	V158	V159	V160	V161	V162	V163	V164	V165	V166	V167	V168	V169	V170	V171	V172	V173	V174	V175	V176	V177	V178	V179	V180	V181	V182	V183	V184	V185	V186	V187	V188	V189	V190	V191	V192	V193	V194	V195	V196	V197	V198	V199	V200	V201	V202	V203	V204	V205	V206	V207	V208	V209	V210	V211	V212	V213	V214	V215	V216	V217	V218	V219	V220	V221	V222	V223	V224	V225	V226	V227	V228	V229	V230	V231	V232	V233	V234	V235	V236	V237	V238	V239	V240	V241	V242	V243	V244	V245	V246	V247	V248	V249	V250	V251	V252	V253	V254	V255	V256	V257	V258	V259	V260	V261	V262	V263	V264	V265	V266	V267	V268	V269	V270	V271	V272	V273	V274	V275	V276	V277	V278	V279	V280	V281	V282	V283	V284	V285	V286	V287	V288	V289	V290	V291	V292	V293	V294	V295	V296	V297	V298	V299	V300	V301	V302	V303	V304	V305	V306	V307	V308	V309	V310	V311	V312	V313	V314	V315	V316	V317	V318	V319	V320	V321	V322	V323	V324	V325	V326	V327	V328	V329	V330	V331	V332	V333	V334	V335	V336	V337	V338	V339	V340	V341	V342	V343	V344	V345	V346	V347	V348	V349	V350	V351	V352	V353	V354	V355	V356	V357	V358	V359	V360	V361	V362	V363	V364	V365	V366	V367	V368	V369	V370	V371	V372	V373	V374	V375	V376	V377	V378	V379	V380	V381	V382	V383	V384	V385	V386	V387	V388	V389	V390	V391	V392	V393	V394	V395	V396	V397	V398	V399	V400	V401	V402	V403	V404	V405	V406	V407	V408	V409	V410	V411	V412	V413	V414	V415	V416	V417	V418	V419	V420	V421	V422	V423	V424	V425	V426	V427	V428	V429	V430	V431	V432	V433	V434	V435	V436	V437	V438	V439	V440	V441	V442	V443	V444	V445	V446	V447	V448	V449	V450	V451	V452	V453	V454	V455	V456	V457	V458	V459	V460	V461	V462	V463	V464	V465	V466	V467	V468	V469	V470	V471	V472	V473	V474	V475	V476	V477	V478	V479	V480	V481	V482	V483	V484	V485	V486	V487	V488	V489	V490	V491	V492	V493	V494	V495	V496	V497	V498	V499	V500	V501	V502	V503	V504	V505	V506	V507	V508	V509	V510	V511	V512	V513	V514	V515	V516	V517	V518	V519	V520	V521	V522	V523	V524	V525	V526	V527	V528	V529	V530	V531	V532	V533	V534	V535	V536	V537	V538	V539	V540	V541	V542	V543	V544	V545	V546	V547	V548	V549	V550	V551	V552	V553	V554	V555	V556	V557	V558	V559	V560	V561	V562	V563	V564	V565	V566	V567	V568	V569	V570	V571	V572	V573	V574	V575	V576	V577	V578	V579	V580	V581	V582	V583	V584	V585	V586	V587	V588	V589	V590	V591	V592	V593	V594	V595	V596	V597	V598	V599	V600	V601	V602	V603	V604	V605	V606	V607	V608	V609	V610	V611	V612	V613	V614	V615	V616	V617	V618	V619	V620	V621	V622	V623	V624	V625	V626	V627	V628	V629	V630	V631	V632	V633	V634	V635	V636	V637	V638	V639	V640	V641	V642	V643	V644	V645	V646	V647	V648	V649	V650	V651	V652	V653	V654	V655	V656	V657	V658	V659	V660	V661	V662	V663	V664	V665	V666	V667	V668	V669	V670	V671	V672	V673	V674	V675	V676	V677	V678	V679	V680	V681	V682	V683	V684	V685	V686	V687	V688	V689	V690	V691	V692	V693	V694	V695	V696	V697	V698	V699	V700	V701	V702	V703	V704	V705	V706	V707	V708	V709	V710	V711	V712	V713	V714	V715	V716	V717	V718	V719	V720	V721	V722	V723	V724	V725	V726	V727	V728	V729	V730	V731	V732	V733	V734	V735	V736	V737	V738	V739	V740	V741	V742	V743	V744	V745	V746	V747	V748	V749	V750	V751	V752	V753	V754	V755	V756	V757	V758	V759	V760	V761	V762	V763	V764	V765	V766	V767	V768	V769	V770	V771	V772	V773	V774	V775	V776	V777	V778	V779	V780	V781	V782	V783	V784	V785	V786	V787	V788	V789	V790	V791	V792	V793	V794	V795	V796	V797	V798	V799	V800	V801	V802	V803	V804	V805	V806	V807	V808	V809	V810	V811	V812	V813	V814	V815	V816	V817	V818	V819	V820	V821	V822	V823	V824	V825	V826	V827	V828	V829	V830	V831	V832	V833	V834	V835	V836	V837	V838	V839	V840	V841	V842	V843	V844	V845	V846	V847	V848	V849	V850	V851	V852	V853	V854	V855	V856	V857	V858	V859	V860	V861	V862	V863	V864	V865	V866	V867	V868	V869	V870	V871	V872	V873	V874	V875	V876	V877	V878	V879	V880	V881	V882	V883	V884	V885	V886	V887	V888	V889	V890	V891	V892	V893	V894	V895	V896	V897	V898	V899	V900	V901	V902	V903	V904	V905	V906	V907	V908	V909	V910	V911	V912	V913	V914	V915	V916	V917	V918	V919	V920	V921	V922	V923	V924	V925	V926	V927	V928	V929	V930	V931	V932	V933	V934	V935	V936	V937	V938	V939	V940	V941	V942	V943	V944	V945	V946	V947	V948	V949	V950	V951	V952	V953	V954	V955	V956	V957	V958	V959	V960	V961	V962	V963	V964	V965	V966	V967	V968	V969	V970	V971	V972	V973	V974	V975	V976	V977	V978	V979	V980	V981	V982	V983	V984	V985	V986	V987	V988	V989	V990	V991	V992	V993	V994	V995	V996	V997	V998	V999	V1000	V1001	V1002	V1003	V1004	V1005	V1006	V1007	V1008	V1009	V1010	V1011	V1012	V1013	V1014	V1015	V1016	V1017	V1018	V1019	V1020	V1021	V1022	V1023	V1024	V1025	V1026	V1027	V1028	V1029	V1030	V1031	V1032	V1033	V1034	V1035	V1036	V1037	V1038	V1039	V1040	V1041	V1042	V1043	V1044	V1045	V1046	V1047	V1048	V1049	V1050	V1051	V1052	V1053	V1054	V1055	V1056	V1057	V1058	V1059	V1060	V1061	V1062	V1063	V1064	V1065	V1066	V1067	V1068	V1069	V1070	V1071	V1072	V1073	V1074	V1075	V1076	V1077	V1078	V1079	V1080	V1081	V1082	V1083	V1084	V1085	V1086	V1087	V1088	V1089	V1090	V1091	V1092	V1093	V1094	V1095	V1096	V1097	V1098	V1099	V1100	V1101	V1102	V1103	V1104	V1105	V1106	V1107	V1108	V1109	V1110	V1111	V1112	V1113	V1114	V1115	V1116	V1117	V1118	V1119	V1120	V1121	V1122	V1123	V1124	V1125	V1126	V1127	V1128	V1129	V1130	V1131	V1132	V1133	V1134	V1135	V1136	V1137	V1138	V1139	V1140	V1141	V1142	V1143	V1144	V1145	V1146	V1147	V1148	V1149	V1150	V1151	V1152	V1153	V1154	V1155	V1156	V1157	V1158	V1159	V1160	V1161	V1162	V1163	V1164	V1165	V1166	V1167	V1168	V1169	V1170	V1171	V1172	V1173	V1174	V1175	V1176	V1177	V1178	V1179	V1180	V1181	V1182	V1183	V1184	V1185	V1186	V1187	V1188	V1189	V1190	V1191	V1192	V1193	V1194	V1195	V1196	V1197	V1198	V1199	V1200	V1201	V1202	V1203	V1204	V1205	V1206	V1207	V1208	V1209	V1210	V1211	V1212	V1213	V1214	V1215	V1216	V1217	V1218	V1219	V1220	V1221	V1222	V1223	V1224	V1225	V1226	V1227	V1228	V1229	V1230	V1231	V1232	V1233	V1234	V1235	V1236	V1237	V1238	V1239	V1240	V1241	V1242	V1243	V1244	V1245	V1246	V1247	V1248	V1249	V1250	V1251	V1252	V1253	V1254	V1255	V1256	V1257	V1258	V1259	V1260	V1261	V1262	V1263	V1264	V1265	V1266	V1267	V1268	V1269	V1270	V1271	V1272	V1273	V1274	V1275	V1276	V1277	V1278	V1279	V1280	V1281	V1282	V1283	V1284	V1285	V1286	V1287	V1288	V1289	V1290	V1291	V1292	V1293	V1294	V1295	V1296	V1297	V1298	V1299	V1300	V1301	V1302	V1303	V1304	V1305	V1306	V1307	V1308	V1309	V1310	V1311	V1312	V1313	V1314	V1315	V1316	V1317	V1318	V1319	V1320	V1321	V1322	V1323	V1324	V1325	V1326	V1327	V1328	V1329	V1330	V1331	V1332	V1333	V1334	V1335	V1336	V1337	V1338	V1339	V1340	V1341	V1342	V1343	V1344	V1345	V1346	V1347	V1348	V1349	V1350	V1351	V1352	V1353	V1354	V1355	V1356	V1357	V1358	V1359	V1360	V1361	V1362	V1363	V1364	V1365	V1366	V1367	V1368	V1369	V1370	V1371	V1372	V1373	V1374	V1375	V1376	V1377	V1378	V1379	V1380	V1381	V1382	V1383	V1384	V1385	V1386	V1387	V1388	V1389	V1390	V1391	V1392	V1393	V1394	V1395	V1396	V1397	V1398	V1399	V1400	V1401	V1402	V1403	V1404	V1405	V1406	V1407	V1408	V1409	V1410	V1411	V1412	V1413	V1414	V1415	V1416	V1417	V1418	V1419	V1420	V1421	V1422	V1423	V1424	V1425	V1426	V1427	V1428	V1429	V1430	V1431	V1432	V1433	V1434	V1435	V1436	V1437	V1438	V1439	V1440	V1441	V1442	V1443	V1444	V1445	V1446	V1447	V1448	V1449	V1450	V1451	V1452	V1453	V1454	V1455	V1456	V1457	V1458	V1459	V1460	V1461	V1462	V1463	V1464	V1465	V1466	V1467	V1468	V1469
------	--------	-------	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

## **Introdução**

As transações financeiras eletrônicas têm crescido exponencialmente, trazendo comodidade para consumidores e empresas. Contudo, esse crescimento também aumentou o risco de fraudes, que causam prejuízos financeiros significativos e afetam a confiança dos usuários nos sistemas de pagamento. A detecção eficiente dessas fraudes é essencial para minimizar perdas e garantir a segurança das operações. Este projeto visa desenvolver uma análise e um modelo preditivo para identificar transações fraudulentas em um conjunto de dados real, utilizando técnicas de ciência de dados e aprendizado de máquina.

## Definição do Problema

O problema central deste projeto é identificar padrões em grandes volumes de dados de transações financeiras que permitam distinguir entre operações legítimas e fraudulentas. Entre os principais desafios estão:

- **Desbalanceamento de Classes:** As fraudes representam uma pequena fração do total de transações, o que dificulta o treinamento de modelos preditivos eficazes.
- **Anonimização dos Dados:** As variáveis do dataset são resultado de uma transformação PCA para proteger a privacidade dos usuários, o que dificulta a interpretação direta.
- **Natureza Dinâmica da Fraude:** Os padrões de fraude podem evoluir com o tempo, exigindo modelos adaptáveis (embora o dataset seja estático, é importante mencionar esse desafio).
- **Volume de Dados:** O grande número de transações exige técnicas eficientes de processamento e análise.

## **Contextualização da Empresa**

O Itaú Unibanco é uma das maiores instituições financeiras do Brasil e da América Latina, oferecendo uma ampla gama de serviços bancários, incluindo cartões de crédito, contas digitais, empréstimos e investimentos. Com milhões de clientes ativos, o banco processa diariamente um grande volume de transações financeiras, tanto em canais físicos quanto digitais.

A detecção de fraudes em transações de cartão de crédito é uma preocupação constante para o Itaú, pois fraudes podem causar prejuízos financeiros significativos, além de afetar a confiança dos clientes na instituição. Fraudes podem ocorrer por meio de clonagem de cartões, uso indevido de dados pessoais ou transações não autorizadas.

Este projeto tem como objetivo desenvolver um modelo preditivo para identificar transações fraudulentas em tempo real, utilizando técnicas de ciência de dados e aprendizado de máquina. Com isso, o Itaú poderá reduzir perdas financeiras, aumentar a segurança das operações e melhorar a experiência dos clientes, bloqueando rapidamente transações suspeitas e evitando transtornos.

## Descrição do Dataset

O conjunto de dados utilizado neste projeto é o Credit Card Fraud Detection, disponibilizado no Kaggle pela Université Libre de Bruxelles (ULB). Ele reúne cerca de 285 mil transações realizadas por titulares de cartões europeus durante o mês de setembro de 2013.

As variáveis presentes no dataset incluem:

- **Time:** representa o tempo, em segundos, decorrido desde a primeira transação registrada até a transação atual.
- **V1 a V28:** são variáveis numéricas obtidas por meio de uma transformação de Análise de Componentes Principais (PCA), aplicada para proteger a privacidade dos usuários, o que dificulta a interpretação direta dessas variáveis.
- **Amount:** valor monetário da transação.
- **Class:** variável alvo que indica se a transação é legítima (0) ou fraudulenta (1).

Algumas características importantes do dataset são o anonimato dos dados, garantido pela transformação PCA, e o forte desbalanceamento entre as classes, já que as fraudes representam uma pequena fração do total de transações. Esse desbalanceamento é um desafio importante para o desenvolvimento de modelos preditivos eficazes.

## **Metodologia Proposta**

### **Definição do Escopo e Planejamento**

O projeto inicia-se com a definição do escopo e o planejamento detalhado das atividades a serem realizadas. Nesta fase, o grupo estabelece os objetivos principais, delimita o problema a ser resolvido, identifica os desafios envolvidos, e organiza um cronograma para guiar o desenvolvimento do trabalho. Essa etapa é fundamental para garantir que todos os integrantes estejam alinhados quanto às metas e responsabilidades.

### **Coleta, Análise Exploratória e Pré-processamento dos Dados**

Em seguida, realiza-se a coleta dos dados, que consiste no download do dataset disponibilizado pela Universidade Livre de Bruxelas (ULB) no Kaggle. Após a obtenção dos dados, é feita uma análise exploratória para compreender a estrutura do dataset, verificar a existência de valores ausentes, analisar a distribuição das variáveis e identificar possíveis outliers. Essa análise é essencial para entender as características dos dados e orientar as etapas seguintes.

Posteriormente, realiza-se o pré-processamento dos dados, que inclui o tratamento de valores ausentes (caso existam), a normalização ou escalonamento das variáveis numéricas, e a aplicação de técnicas para lidar com o desbalanceamento da classe alvo, como oversampling, undersampling ou SMOTE. Essa etapa prepara os dados para que os modelos de machine learning possam ser treinados de forma eficaz.

### **Modelagem, Treinamento e Avaliação**

Com os dados preparados, a próxima fase envolve a modelagem e o treinamento dos algoritmos de aprendizado de máquina. Serão utilizados diversos modelos, como Regressão Logística, Árvores de Decisão, Random Forest, Gradient Boosting, Support Vector Machines e Redes Neurais, para identificar transações fraudulentas. Após o treinamento, os modelos serão avaliados utilizando métricas apropriadas para problemas desbalanceados, como Precisão, Recall, F1-Score e AUC-ROC. Também serão realizados ajustes nos hiperparâmetros para otimizar o desempenho dos modelos.

## **Documentação, Apresentação e Repositório**

Por fim, o projeto será documentado detalhadamente, incluindo a descrição das metodologias utilizadas, os resultados obtidos, as limitações encontradas e as recomendações para trabalhos futuros. A apresentação dos resultados será preparada para compartilhar as conclusões com a equipe e o professor.

Além disso, todo o código-fonte, scripts e documentos relacionados ao projeto serão organizados e versionados em um repositório online, como o GitHub ou GitLab. O uso do repositório garante o controle de versões, facilita a colaboração entre os integrantes e permite o acesso remoto ao material do projeto.



## Cronograma

Etapa	Atividade	Responsável	Início	Término	Entrega / Milestone
1	Definição do escopo e tema do projeto	Todos os membros	30/08/2025	30/08/2025	Tema definido
1	Escolha e validação do dataset	Déborah	30/08/2025	03/09/2025	Dataset aprovado
1	Organização do cronograma e divisão de tarefas	Déborah	31/09/2025	04/09/2025	Cronograma fechado
1	<b>Entrega da Etapa 1</b>	<b>Todos os membros</b>		<b>04/09/2025</b>	<b>A1 – Aplicando Conhecimento (N1)</b>
2	Coleta do dataset e análise exploratória	Lucas	05/09/2025	12/09/2025	Relatório exploratório
2	Pré-processamento e balanceamento dos dados	Lucas	13/09/2025	20/09/2025	Dataset tratado
2	Redação parcial e revisão técnica	Luiz	01/10/2025	15/10/2025	Documento intermediário
2	<b>Entrega da Etapa 2</b>	<b>Déborah</b>		<b>16/10/2025</b>	<b>A2 – Aplicando Conhecimento (N1)</b>
3	Modelagem e treinamento dos algoritmos	Diógenes	17/10/2025	25/10/2025	Modelos treinados
3	Avaliação dos modelos e ajustes	Diógenes	26/10/2025	03/11/2025	Relatório de desempenho
3	Geração de gráficos e visualizações	Déborah	04/11/2025	10/11/2025	Material visual pronto
3	<b>Entrega da Etapa 3</b>	<b>Déborah</b>		<b>13/11/2025</b>	<b>A3 – Atividades</b>
4	Documentação final e organização do repositório	Luiz	14/11/2025	23/11/2025	Relatório final concluído
4	Preparação da apresentação e vídeo final	Todos os membros	24/11/2025	26/11/2025	Apresentação pronta
4	<b>Entrega da Etapa 4</b>	<b>Todos os membros</b>		<b>27/11/2025</b>	<b>A4 – Projeto Final (PF)</b>

## Repositório do Projeto

O código-fonte, scripts e documentos relacionados a este projeto estão organizados no repositório GitHub disponível em: [https://github.com/httpsdebs/Projeto\\_Aplicado\\_I](https://github.com/httpsdebs/Projeto_Aplicado_I)

O repositório está estruturado em pastas para dados, notebooks, scripts e documentação, contendo um arquivo README com a descrição do projeto, objetivos e nomes dos integrantes.

## Referências

BRUSSELS, Université Libre de. *Credit Card Fraud Detection Dataset*. Disponível em: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>. Acesso em: 04 set. 2025.

ITAÚ UNIBANCO. *Site institucional*. Disponível em: <https://www.itaubank.com.br>. Acesso em: 04 set. 2025.

SANTOS, João; SILVA, Maria. *Aprendizado de Máquina aplicado à detecção de fraudes em cartões de crédito*. São Paulo: Editora Acadêmica, 2023.