QUANTUM WOLF

DATA INTELLIGENCE RESEARCH LAB

DATE:14-02-2025 TO 16-02-2025

NAME: ABDUL ZUHAIL M

Fraud Detection System for Loan Transactions

1. Problem Statement

The financial industry faces increasing challenges in detecting fraudulent loan transactions. Fraudulent activities include misrepresentation of financial data, false loan purposes, and deliberate repayment defaults, leading to significant financial losses. Traditional fraud detection methods are often:

- **Manual** – Requiring human intervention, making them slow and inefficient.
- **Error-Prone** – Increasing the chances of oversight and false classifications.
- **Time-Consuming** – Requiring extensive effort to analyze and verify transactions.

Why Is an Automated Fraud Detection System Needed?

A robust fraud detection system should:

- **Accurately detect fraudulent loan transactions.**
- **Provide real-time alerts** for high-risk transactions to prevent further financial losses.
- **Visualize fraud trends and patterns** for better decision-making and risk management.

---

2. Solution

To address these challenges, we developed a **Fraud Detection System using KNIME**, a powerful data analytics platform. This system utilizes **Machine Learning (Random Forest algorithm) and rule-based filtering** to:

- **Preprocess and clean loan transaction data.**
- **Train a fraud detection model** to classify transactions as fraudulent or genuine.
- **Predict fraud risk scores** for new transactions.
- **Generate real-time alerts** for suspicious transactions.
- **Provide an interactive dashboard** for monitoring and analysis.

By implementing this system, financial institutions can detect fraud more efficiently and make informed decisions based on real-time insights.

---

3. Working Process

The workflow of the fraud detection system is divided into six main steps:

Step 1: Data Ingestion

Objective: Load and clean the dataset to remove inconsistencies.

Process:

- **CSV Reader Node:** Loads the dataset from a file (e.g., loan_fraud_dataset.csv).
- **Column Filter Node:** Removes unnecessary or irrelevant columns, keeping only those essential for fraud detection (e.g., loan amount, loan purpose, credit score, repayment status).

Output: A cleaned dataset that contains only relevant columns, making the data more structured and easier to analyze.

Expected Output:

Filtered table (Table)

Rows: 100 | Columns: 11

#	RowID	Industry (String)	Loan Amount... Number (integer)	Loan Purpose (String)	Annual Reve... Number (integer)	Cash Flow (U... Number (integer)	Loan Applica... Number (integer)	Loan Applica... Number (integer)	Actual Reven... Number (integer)	Actual Profit ... Number (integer)	Loan Repaym... (String)	Fraud Indicator Number (integer)
1	Row0	Manufacturing	64541	Equipment Purchas	870817	222438	830400	218425	879392	196016	Delayed	0
2	Row1	Healthcare	394798	Expansion	703326	185934	751033	167635	744821	229682	Delayed	0
3	Row2	Healthcare	265949	Debt Consolidation	459463	86666	411878	73702	438395	68561	Delayed	0
4	Row3	Healthcare	379951	Working Capital	712442	199802	699408	184896	657033	237685	Defaulted	1
5	Row4	Agriculture	115419	Working Capital	123352	247433	130577	252789	119913	223138	Paid	0
6	Row5	Healthcare	238264	Working Capital	478604	-34418	504739	-26347	553442	-25704	Defaulted	1
7	Row6	Manufacturing	490823	Equipment Purchas	679758	170088	714487	165523	621599	174378	Defaulted	0
8	Row7	Agriculture	215113	Expansion	239613	162432	201872	155318	142802	140596	Defaulted	1
9	Row8	Agriculture	382711	Debt Consolidation	127658	284378	94029	296013	33523	292465	Delayed	0
10	Row9	Agriculture	383391	Working Capital	756688	89173	770495	80296	751855	110031	Paid	0
11	Row10	Healthcare	373982	Expansion	136244	75838	176062	78317	183373	67035	Delayed	0
12	Row11	Manufacturing	114091	Debt Consolidation	730128	101962	724952	93528	672858	139952	Paid	0
13	Row12	Healthcare	220920	Equipment Purchas	791246	158266	757507	176374	716184	120876	Defaulted	1
14	Row13	Agriculture	353907	Equipment Purchas	461882	125791	434091	118934	468823	130665	Paid	0
15	Row14	Retail	462691	Working Capital	655051	-1444	657043	11038	734503	-19689	Paid	0
16	Row15	Healthcare	293031	Working Capital	898352	-19989	931909	-13177	811063	865	Defaulted	1
17	Row16	Retail	299031	Expansion	987543	246875	962735	265889	982620	262614	Paid	0
18	Row17	Healthcare	158978	Equipment Purchas	454075	261077	439772	267305	484452	286706	Paid	0
19	Row18	Retail	353146	Working Capital	743900	17442	719543	24825	743968	3457	Paid	0
20	Row19	Agriculture	408561	Working Capital	846369	126810	839775	138475	933757	176737	Paid	0
21	Row20	Technology	442886	Expansion	476472	50833	506849	60358	551271	45990	Delayed	0
22	Row21	Manufacturing	75135	Debt Consolidation	850324	-14744	879885	-32185	842760	-4144	Paid	0
23	Row22	Manufacturing	112484	Expansion	188766	67832	147039	73146	133424	49388	Delayed	0
24	Row23	Retail	470458	Debt Consolidation	542521	46182	515600	27401	617559	88011	Paid	0
25	Row24	Technology	246243	Equipment Purchas	374020	127083	386342	147077	345345	115133	Delayed	0
26	Row25	Agriculture	115208	Debt Consolidation	372044	29381	357455	43574	300912	65525	Delayed	0
27	Row26	Technology	213755	Equipment Purchas	219052	90368	228898	107183	267068	91954	Defaulted	1
28	Row27	Manufacturing	284518	Working Capital	344150	34862	305756	17022	271203	73515	Paid	0
29	Row28	Manufacturing	287099	Expansion	545151	298434	526953	297063	621794	331513	Delayed	0
30	Row29	Retail	499104	Working Capital	113899	20606	90787	34402	97057	27880	Defaulted	0
31	Row30	Manufacturing	94102	Equipment Purchas	656779	66527	701810	55624	668363	91170	Paid	0
32	Row31	Technology	202263	Debt Consolidation	583821	281740	588966	286546	494554	324570	Delayed	0

Step 2: Data Preprocessing & Transformation

Objective: Prepare the dataset for model training by handling missing values and transforming categorical data.

Process:

- **Missing Value Handler Node:** Fills in missing data for better model performance:
  - **Numerical columns** (e.g., loan amount, annual revenue) – Filled with mean or median values.
  - **Categorical columns** (e.g., loan type, industry) – Filled with the most frequently occurring value.
- **One Hot Encoder Node:** Converts categorical data (e.g., industry type, loan purpose) into numerical format, making it suitable for machine learning models.

- **Normalizer Node:** Scales numerical features (e.g., loan amount, annual revenue) to ensure consistent data ranges, improving model accuracy.

Output: A fully processed and structured dataset, ready for training the machine learning model.

Expected Output:

Normalized table (Table)

Rows: 100 | Columns: 23

#	RowID	Industry String	Loan Am... Number (dou...	Loan Pur... String	Annual R... Number (dou...	Cash Flo... Number (dou...	Loan App... Number (dou...	Loan App... Number (dou...	Actual Re... Number (dou...	Actual Pr... Number (dou...	Loan Rep... String	Fraud Ind... Number (dou...	Manufact... Number (inte...	Healthcare Number (inte...	Agricu Number
1	Row0	Manufacturing	0.002	Equipment Purc	0.854	0.777	0.836	0.78	0.842	0.668	Delayed	0	1	0	0
2	Row1	Healthcare	0.761	Expansion	0.665	0.671	0.746	0.638	0.708	0.75	Delayed	0	0	1	0
3	Row2	Healthcare	0.465	Debt Consolida	0.39	0.384	0.363	0.375	0.403	0.356	Delayed	0	0	1	0
4	Row3	Healthcare	0.726	Working Capita	0.676	0.711	0.688	0.686	0.62	0.77	Defaulted	1	0	1	0
5	Row4	Agriculture	0.119	Working Capita	0.011	0.849	0.045	0.876	0.086	0.734	Paid	0	0	0	1
6	Row5	Healthcare	0.401	Working Capita	0.412	0.034	0.468	0.095	0.517	0.126	Defaulted	1	0	1	0
7	Row6	Manufacturing	0.981	Equipment Purc	0.639	0.625	0.705	0.632	0.585	0.615	Defaulted	0	1	0	0
8	Row7	Agriculture	0.348	Expansion	0.142	0.603	0.126	0.603	0.109	0.532	Defaulted	1	0	0	1
9	Row8	Agriculture	0.733	Debt Consolida	0.016	0.956	0.004	0.997	0	0.904	Delayed	0	0	0	1
10	Row9	Agriculture	0.734	Working Capita	0.726	0.391	0.768	0.393	0.715	0.458	Paid	0	0	0	1
11	Row10	Healthcare	0.713	Expansion	0.025	0.353	0.096	0.388	0.149	0.353	Delayed	0	0	1	0
12	Row11	Manufacturing	0.116	Debt Consolida	0.696	0.428	0.717	0.43	0.636	0.531	Paid	0	1	0	0
13	Row12	Healthcare	0.361	Equipment Purc	0.765	0.591	0.754	0.662	0.679	0.484	Defaulted	1	0	1	0
14	Row13	Agriculture	0.667	Equipment Purc	0.393	0.497	0.388	0.501	0.433	0.508	Paid	0	0	0	1
15	Row14	Retail	0.916	Working Capita	0.611	0.129	0.64	0.199	0.698	0.14	Paid	0	0	0	0

Step 3: Model Training & Evaluation

Objective: Train a fraud detection model and assess its performance.

Process:

- **Partitioning Node:** Splits the dataset into:
  - **Training Set (70%)** – Used to train the model.
  - **Testing Set (30%)** – Used to evaluate model performance.
- **Random Forest Learner Node:** Trains the machine learning model using key features such as:
  - Loan Amount
  - Annual Revenue
  - Loan Repayment Status

- Credit Score

- **Scorer Node:** Evaluates model performance using:
  - **Accuracy** – Measures overall correctness of the model.
  - **Precision** – Measures the percentage of correctly identified fraudulent transactions.
  - **Recall** – Measures the percentage of actual fraud cases detected.
  - **F1-Score** – A balance between precision and recall, ensuring a robust model.

Output: A trained and validated fraud detection model, capable of identifying fraudulent transactions.

Expected Output:

Confusion matrix (Table)

Rows: 3 | Columns: 3

#	RowID	Delayed Number (integer)	Defaulted Number (integer)	Paid Number (integer)
1	Delaye	8	11	3
2	Default	9	7	9
3	Paid	8	7	8

Accuracy statistics (Table)

Rows: 4 | Columns: 11

#	RowID	TruePositives Number (integer)	FalsePositives Number (integer)	TrueNegatives Number (integer)	FalseNegatives Number (integer)	Recall Number (double)	Precision Number (double)	Sensitivity Number (double)	Specificity Number (double)	F-measure Number (double)	Accuracy Number (double)	Cohen's kappa Number (double)
1	Delaye	8	17	31	14	0.364	0.32	0.364	0.646	0.34	?	?
2	Default	7	18	27	18	0.28	0.28	0.28	0.6	0.28	?	?
3	Paid	8	12	35	15	0.348	0.4	0.348	0.745	0.372	?	?
4	Overall	?	?	?	?	?	?	?	?	?	0.329	-0.008

Step 4: Fraud Prediction on New Transactions

Objective: Predict fraud risk for new loan applications.

Process:

- **Random Forest Predictor Node:** Uses the trained model to classify new transactions as fraudulent or genuine.

Expected Output:

Prediction output (table)

Rows: 30 | Columns: 25

#	RowID	Industry	Loan Am...	Loan Pur...	Annual R...	Cash Flo...	Loan App...	Loan App...	Actual Re...	Actual Pr...	Loan Rep...	Fraud Ind...	Manufact...	Healthcare
1	Row0	Manufacturing	0.002	Equipment Purc	0.854	0.777	0.836	0.78	0.842	0.668	Delayed	0	1	0
2	Row3	Healthcare	0.726	Working Capita	0.676	0.711	0.688	0.686	0.62	0.77	Defaulted	1	0	1
3	Row5	Healthcare	0.401	Working Capita	0.412	0.034	0.468	0.095	0.517	0.126	Defaulted	1	0	1
4	Row11	Manufacturing	0.116	Debt Consolida	0.696	0.428	0.717	0.43	0.636	0.531	Paid	0	1	0
5	Row13	Agriculture	0.667	Equipment Purc	0.393	0.497	0.388	0.501	0.433	0.508	Paid	0	0	0
6	Row17	Healthcare	0.219	Equipment Purc	0.384	0.889	0.394	0.917	0.449	0.89	Paid	0	0	1
7	Row20	Technology	0.871	Expansion	0.409	0.28	0.47	0.337	0.515	0.301	Delayed	0	0	0
8	Row21	Manufacturing	0.027	Debt Consolida	0.831	0.091	0.892	0.078	0.805	0.179	Paid	0	1	0
9	Row25	Agriculture	0.119	Debt Consolida	0.291	0.218	0.301	0.291	0.266	0.349	Delayed	0	0	0
10	Row28	Manufacturing	0.513	Expansion	0.487	0.997	0.493	1	0.585	0.999	Delayed	0	1	0
11	Row30	Manufacturing	0.07	Equipment Purc	0.613	0.326	0.691	0.324	0.632	0.412	Paid	0	1	0
12	Row38	Retail	0.999	Working Capita	0.992	0.181	0.974	0.223	1	0.109	Paid	0	0	0
13	Row39	Retail	0.985	Working Capita	0.044	0.377	0.034	0.383	0.196	0.355	Defaulted	1	0	0
14	Row40	Retail	0.992	Debt Consolida	0.027	0.904	0.049	0.907	0.146	0.83	Paid	0	0	0
15	Row43	Agriculture	0.745	Expansion	0.106	0.482	0.079	0.458	0.25	0.57	Defaulted	0	0	0
16	Row45	Manufacturing	0.02	Working Capita	0.834	0.54	0.817	0.559	0.897	0.432	Paid	0	1	0
17	Row46	Retail	0.447	Equipment Purc	0.38	0.588	0.377	0.576	0.509	0.677	Paid	0	0	0
18	Row52	Agriculture	0.695	Debt Consolida	0.115	0.826	0.148	0.86	0.228	0.655	Defaulted	0	0	0
19	Row55	Manufacturing	0.061	Expansion	0.984	0.352	0.958	0.329	0.921	0.494	Defaulted	1	1	0
20	Row57	Manufacturing	0.849	Expansion	0.796	0.533	0.872	0.525	0.835	0.466	Paid	0	1	0
21	Row61	Technology	0.812	Working Capita	0.487	0.109	0.492	0.134	0.498	0.226	Paid	0	0	0
22	Row62	Healthcare	0.768	Expansion	0.134	0.841	0.125	0.907	0.149	0.698	Paid	0	0	1
23	Row65	Healthcare	0.295	Working Capita	0.103	0.566	0.164	0.556	0.216	0.477	Delayed	0	0	1
24	Row71	Technology	0.937	Expansion	0.342	0.926	0.382	0.909	0.476	0.825	Delayed	0	0	0
25	Row81	Retail	0.423	Equipment Purc	0.747	0.61	0.747	0.62	0.824	0.549	Paid	0	0	0
26	Row82	Retail	0.79	Equipment Purc	0.738	0.702	0.723	0.753	0.721	0.616	Paid	0	0	0
27	Row85	Agriculture	0.805	Equipment Purc	0.059	0.065	0.055	0.13	0.061	0.188	Paid	0	0	0
28	Row91	Manufacturing	0.642	Expansion	0.87	0.855	0.868	0.858	0.861	0.736	Delayed	0	1	0
29	Row93	Technology	0.592	Equipment Purc	0.699	0.494	0.718	0.569	0.601	0.499	Paid	0	0	0
30	Row95	Retail	0.751	Debt Consolida	0.535	0.955	0.528	0.911	0.575	0.93	Delayed	0	0	0

#	Agriculture	Retail	Technolo...	Equipme...	Expansion	Debt Con...	Working ...	Delayed	Defaulted	Paid	Predictio...	Predictio...
1	[illegible]	[illegible]	0	1	0	0	0	1	0	0	Defaulted	0.43
2	0	0	0	0	0	0	1	0	1	0	Paid	0.35
3	0	0	0	0	0	0	1	0	1	0	Delayed	0.41
4	0	0	0	0	0	1	0	0	0	1	Defaulted	0.37
5	1	0	0	1	0	0	0	0	0	1	Defaulted	0.37
6	0	0	0	1	0	0	0	0	0	1	Delayed	0.59
7	0	0	1	0	1	0	0	1	0	0	Paid	0.39
8	0	0	0	0	0	1	0	0	0	1	Defaulted	0.44
9	1	0	0	0	0	1	0	1	0	0	Defaulted	0.4
10	0	0	0	0	1	0	0	1	0	0	Delayed	0.68
11	0	0	0	1	0	0	0	0	0	1	Defaulted	0.42
12	0	1	0	0	0	0	1	0	0	1	Paid	0.5
13	0	1	0	0	0	0	1	0	1	0	Defaulted	0.48
14	0	1	0	0	0	1	0	0	0	1	Delayed	0.55
15	1	0	0	0	1	0	0	0	1	0	Delayed	0.38
16	0	0	0	0	0	0	1	0	0	1	Paid	0.39
17	0	1	0	1	0	0	0	0	0	1	Defaulted	0.47
18	1	0	0	0	0	1	0	0	1	0	Defaulted	0.48
19	0	0	0	0	1	0	0	0	1	0	Paid	0.58
20	0	0	0	0	1	0	0	0	0	1	Defaulted	0.42
21	0	0	1	0	0	0	1	0	0	1	Defaulted	0.44
22	0	0	0	0	1	0	0	0	0	1	Delayed	0.52
23	0	0	0	0	0	0	1	1	0	0	Defaulted	0.45
24	0	0	1	0	1	0	0	1	0	0	Delayed	0.68
25	0	1	0	1	0	0	0	0	0	1	Defaulted	0.5
26	0	1	0	1	0	0	0	0	0	1	Defaulted	0.57
27	1	0	0	1	0	0	0	0	0	1	Paid	0.45
28	0	0	0	0	1	0	0	1	0	0	Delayed	0.45
29	0	0	1	1	0	0	0	0	0	1	Defaulted	0.45
30	0	1	0	0	0	1	0	1	0	0	Delayed	0.56

- **Math Formula Node:** Computes a risk score for each transaction using the formula:

  (1 - $Prediction (Loan Repayment Status) (Confidence)$) * $Loan Amount (USD)$

- **Rule-Based Row Filter Node:** Flags transactions as **high-risk if**

  $Risk Score$ > 0.2 => TRUE

  TRUE => FALSE

Output:

- Each new transaction is classified as **genuine or fraudulent**.
- High-risk transactions are flagged for further review.

Expected Output:

Filtered (Table)

Rows: 20 | Columns: 26

#	RowID	Industry (String)	Loan Am... (Number (dou...)	Loan Pur... (String)	Annual R... (Number (dou...)	Cash Flo... (Number (dou...)	Loan App... (Number (dou...)	Loan App... (Number (dou...)	Actual Re... (Number (dou...)	Actual Pr... (Number (dou...)	Loan Rep... (String)	Fraud Ind... (Number (dou...)	Manufact... (Number (inte...)	Healthcare (Number (inte...)	Agricultu (Number (int)
1	Row3	Healthcare	0.726	Working Capita	0.676	0.711	0.688	0.686	0.62	0.77	Defaulted	1	0	1	0
2	Row5	Healthcare	0.401	Working Capita	0.412	0.034	0.468	0.095	0.517	0.126	Defaulted	1	0	1	0
3	Row13	Agriculture	0.667	Equipment Purc	0.393	0.497	0.388	0.501	0.433	0.508	Paid	0	0	0	1
4	Row20	Technology	0.871	Expansion	0.409	0.28	0.47	0.337	0.515	0.301	Delayed	0	0	0	0
5	Row38	Retail	0.999	Working Capita	0.992	0.181	0.974	0.223	1	0.109	Paid	0	0	0	0
6	Row39	Retail	0.985	Working Capita	0.044	0.377	0.034	0.383	0.196	0.355	Defaulted	1	0	0	0
7	Row40	Retail	0.992	Debt Consolida	0.027	0.904	0.049	0.907	0.146	0.83	Paid	0	0	0	0
8	Row43	Agriculture	0.745	Expansion	0.106	0.482	0.079	0.458	0.25	0.57	Defaulted	0	0	0	1
9	Row46	Retail	0.447	Equipment Purc	0.38	0.588	0.377	0.576	0.509	0.677	Paid	0	0	0	0
10	Row52	Agriculture	0.695	Debt Consolida	0.115	0.826	0.148	0.86	0.228	0.655	Defaulted	0	0	0	1
11	Row57	Manufacturing	0.849	Expansion	0.796	0.533	0.872	0.525	0.835	0.466	Paid	0	1	0	0
12	Row61	Technology	0.812	Working Capita	0.487	0.109	0.492	0.134	0.498	0.226	Paid	0	0	0	0
13	Row62	Healthcare	0.768	Expansion	0.134	0.841	0.125	0.907	0.149	0.698	Paid	0	0	1	0
14	Row71	Technology	0.937	Expansion	0.342	0.926	0.382	0.909	0.476	0.825	Delayed	0	0	0	0
15	Row81	Retail	0.423	Equipment Purc	0.747	0.61	0.747	0.62	0.824	0.549	Paid	0	0	0	0
16	Row82	Retail	0.79	Equipment Purc	0.738	0.702	0.723	0.753	0.721	0.616	Paid	0	0	0	0
17	Row85	Agriculture	0.805	Equipment Purc	0.059	0.065	0.055	0.13	0.061	0.188	Paid	0	0	0	1
18	Row91	Manufacturing	0.642	Expansion	0.87	0.855	0.868	0.858	0.861	0.736	Delayed	0	1	0	0
19	Row93	Technology	0.592	Equipment Purc	0.699	0.494	0.718	0.569	0.601	0.499	Paid	0	0	0	0
20	Row95	Retail	0.751	Debt Consolida	0.535	0.955	0.528	0.911	0.575	0.93	Delayed	0	0	0	0

Step 5: Fraud Alerts & Storage

Objective: Notify stakeholders about potential fraud and store fraud cases for future analysis.

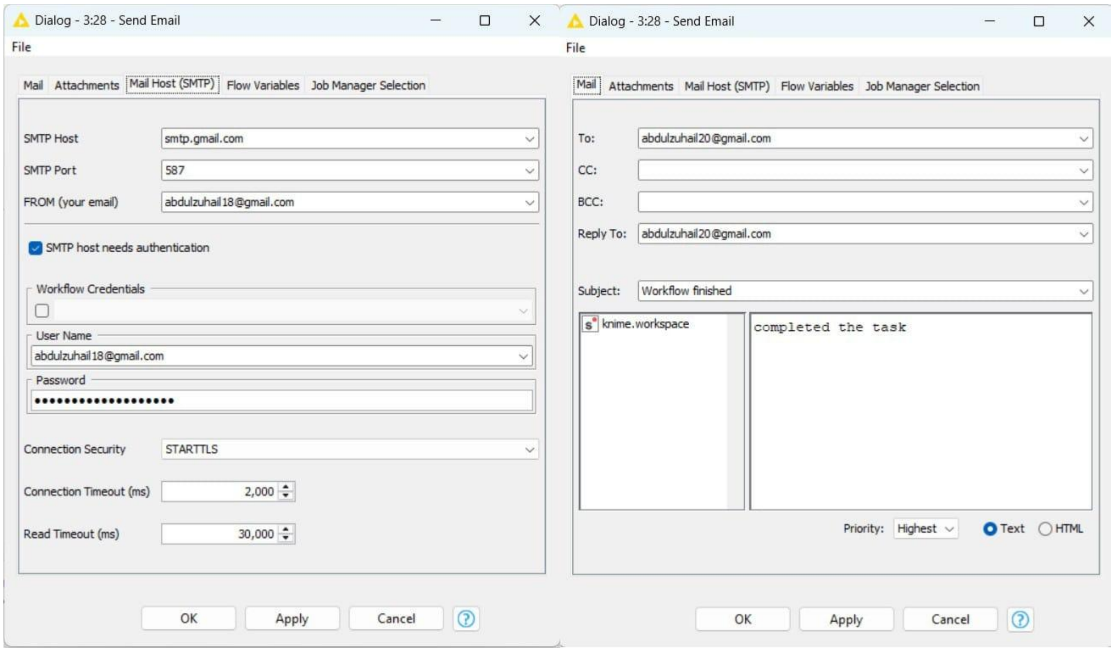
Process:

- **Rule-Based Row Filter Node:** Extracts transactions flagged as fraudulent.
- **Send Email Node:** Sends alerts with transaction details to stakeholders via email.
- **CSV Writer Node:** Saves the fraud detection results for future auditing.

Output:

- Real-time email alerts for high-risk transactions.
- A stored dataset for further investigation and improvements.

Expected Output:



Step 6: Dashboard & Visualization

Objective: Provide an interactive dashboard for fraud monitoring and trend analysis.

Process:

- **Table View Node:** Displays a structured table of fraud cases.
- **Visualization Nodes:** Generate insights through:
  - **Bar Chart:** Shows fraud trends by industry and loan purpose.
  - **Pie Chart:** Illustrates the percentage of fraudulent vs. genuine transactions.
  - **Scatter Plot:** Plots fraud risk scores against loan amounts to identify patterns.

Output: A user-friendly dashboard that helps financial analysts monitor fraud trends and take proactive measures.

Expected Output:

Table View

Interactive View: Table View

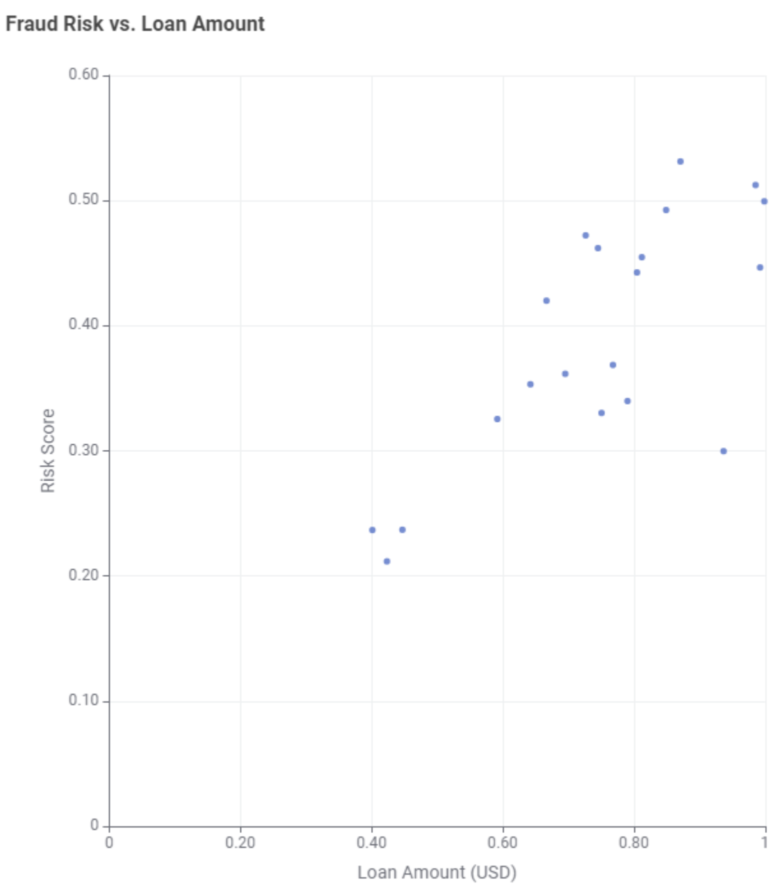
Table View

Rows: 20 | Columns: 26

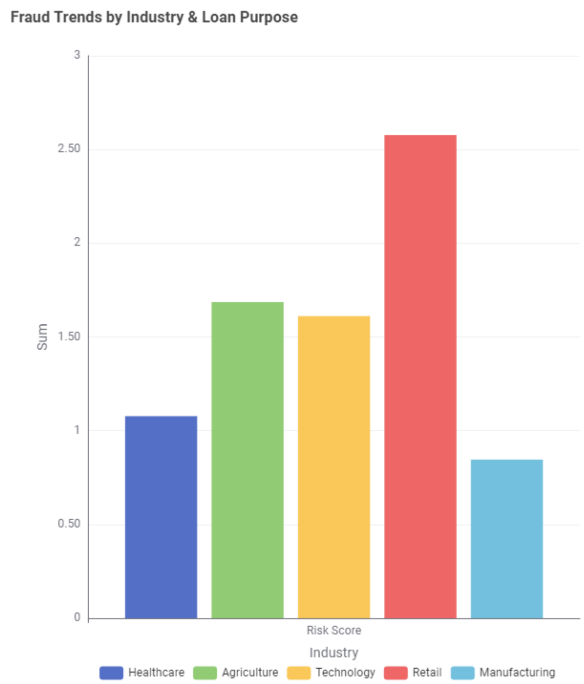
RowID	Industry	Loan Am...	Loan Pur...	Annual R...	Cash Flo...	Loan App...	Loan App...	Actual Re...	Actual Pr...	Loan Rep...	Fraud Ind...	Manufact...	Healthcare	Agriculture
Row3	Healthcare	0.726	Working Capita	0.676	0.711	0.688	0.686	0.62	0.77	Defaulted	1	0	1	0
Row5	Healthcare	0.401	Working Capita	0.412	0.034	0.468	0.096	0.517	0.126	Defaulted	1	0	1	0
Row13	Agriculture	0.667	Equipment Purc	0.393	0.497	0.388	0.501	0.433	0.508	Paid	0	0	0	1
Row20	Technology	0.871	Expansion	0.409	0.28	0.47	0.337	0.515	0.301	Delayed	0	0	0	0
Row38	Retail	0.999	Working Capita	0.992	0.181	0.974	0.223	1	0.109	Paid	0	0	0	0
Row39	Retail	0.985	Working Capita	0.044	0.377	0.034	0.383	0.196	0.355	Defaulted	1	0	0	0
Row40	Retail	0.992	Debt Consolida	0.027	0.904	0.049	0.907	0.146	0.83	Paid	0	0	0	0
Row43	Agriculture	0.745	Expansion	0.106	0.482	0.079	0.458	0.25	0.57	Defaulted	0	0	0	1
Row46	Retail	0.447	Equipment Purc	0.38	0.588	0.377	0.576	0.509	0.677	Paid	0	0	0	0
Row52	Agriculture	0.695	Debt Consolida	0.115	0.826	0.148	0.86	0.228	0.655	Defaulted	0	0	0	1
Row57	Manufacturing	0.849	Expansion	0.796	0.533	0.872	0.525	0.835	0.466	Paid	0	1	0	0
Row61	Technology	0.812	Working Capita	0.487	0.109	0.492	0.134	0.498	0.226	Paid	0	0	0	0
Row62	Healthcare	0.768	Expansion	0.134	0.841	0.125	0.907	0.149	0.698	Paid	0	0	1	0
Row71	Technology	0.937	Expansion	0.342	0.926	0.382	0.909	0.476	0.825	Delayed	0	0	0	0
Row81	Retail	0.423	Equipment Purc	0.747	0.61	0.747	0.62	0.824	0.549	Paid	0	0	0	0
Row82	Retail	0.79	Equipment Purc	0.738	0.702	0.723	0.753	0.721	0.616	Paid	0	0	0	0
Row85	Agriculture	0.805	Equipment Purc	0.059	0.065	0.055	0.13	0.061	0.188	Paid	0	0	0	1
Row91	Manufacturing	0.642	Expansion	0.87	0.855	0.868	0.858	0.861	0.736	Delayed	0	1	0	0

Retail	Technolo...	Equipme...	Expansion	Debt Con...	Working ...	Delayed	Defaulted	Paid	Predictio...	Predictio...	Risk Score
0	0	0	0	0	1	0	1	0	Paid	0.35	0.472
0	0	0	0	0	1	0	1	0	Delayed	0.41	0.237
0	0	1	0	0	0	0	0	1	Defaulted	0.37	0.42
0	1	0	1	0	0	1	0	0	Paid	0.39	0.531
1	0	0	0	0	1	0	0	1	Paid	0.5	0.499
1	0	0	0	0	1	0	1	0	Defaulted	0.48	0.512
1	0	0	0	1	0	0	0	1	Delayed	0.55	0.447
0	0	0	1	0	0	0	1	0	Delayed	0.38	0.462
1	0	1	0	0	0	0	0	1	Defaulted	0.47	0.237
0	0	0	0	1	0	0	1	0	Defaulted	0.48	0.362
0	0	0	1	0	0	0	0	1	Defaulted	0.42	0.492
0	1	0	0	0	1	0	0	1	Defaulted	0.44	0.455
0	0	0	1	0	0	0	0	1	Delayed	0.52	0.369
0	1	0	1	0	0	1	0	0	Delayed	0.68	0.3
1	0	1	0	0	0	0	0	1	Defaulted	0.5	0.212
1	0	1	0	0	0	0	0	1	Defaulted	0.57	0.34
0	0	1	0	0	0	0	0	1	Paid	0.45	0.443
0	0	0	1	0	0	1	0	0	Delayed	0.45	0.353

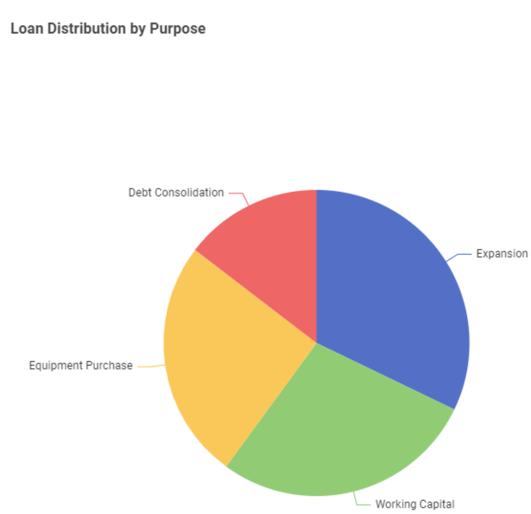
Scatter Plot



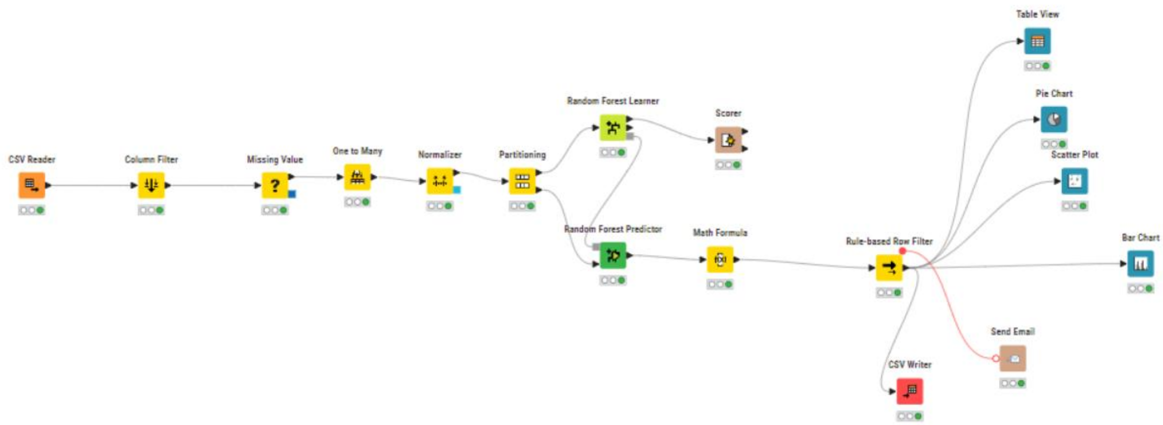
Bar Chart



Pie Chart



Workflow Diagram



Workflow Path Overview

1. **CSV Reader –** Reads data from a CSV file.
2. **Column Filter –** Selects relevant columns.
3. **Missing Value Handling** – Manages missing data**.**
4. **One to Many –** Splits single records into multiple if needed.
5. **Normalizer –** Standardizes data for analysis**.**
6. **Partitioning –** Splits data into training and testing sets.
7. **Random Forest Learner –** Trains a Random Forest model.
8. **Random Forest Predictor –** Makes predictions on test data.
9. **Scorer –** Evaluates model performance.
10. **Math Formulas –** Applies necessary calculations**.**
11. **Rule-based Row Filter –** Filters data based on rules.
12. **Output Steps:**

- **Table View –** Displays processed data.

- **Pie Chart, Scatter Plot, Bar Chart –** Data visualizations.
- **CSV Writer** – Saves processed data to a file.
- **Send Email –** Notifies users with results.

---

4. Final Output

The system successfully provides:

- A **trained fraud detection model** with optimized accuracy.
- A **classification system** that assigns risk scores to transactions.
- **Real-time email alerts** for high-risk transactions.
- An **interactive dashboard** to visualize fraud patterns and trends.

---

5. Challenges Faced

Data Quality Issues:

- Missing values in critical columns required careful imputation.
- Inconsistent data formats needed to be standardized.

Model Performance:

- Early models struggled with low recall, missing some fraudulent cases.
- Balancing **precision and recall** was difficult, requiring multiple iterations.

Dashboard Design:

- Making the dashboard **user-friendly and visually intuitive** required multiple refinements.

AI Assistance Challenges:

- ChatGPT provided multiple suggestions at the start, making it easier to proceed.
- However, when specific issues arose, the provided solutions were often varied, sometimes requiring us to restart from the beginning to align everything properly.

---

6. Conclusion

The **Fraud Detection System for Loan Transactions** automates fraud identification by combining machine learning and rule-based filtering. The system successfully:

- **Identifies fraudulent transactions** with high accuracy.
- **Generates real-time alerts** for suspicious transactions.
- **Provides an interactive dashboard** for fraud monitoring and analysis.

Despite challenges in data quality and model tuning, the system proves to be a **robust and scalable solution** for detecting fraud in the financial industry. Future enhancements could include:

- **Real-time data streaming** for instant fraud detection.
- **Advanced machine learning models** such as deep learning to improve accuracy.

This report provides a **comprehensive guide** on building a fraud detection system using KNIME, demonstrating its effectiveness in combating financial fraud.