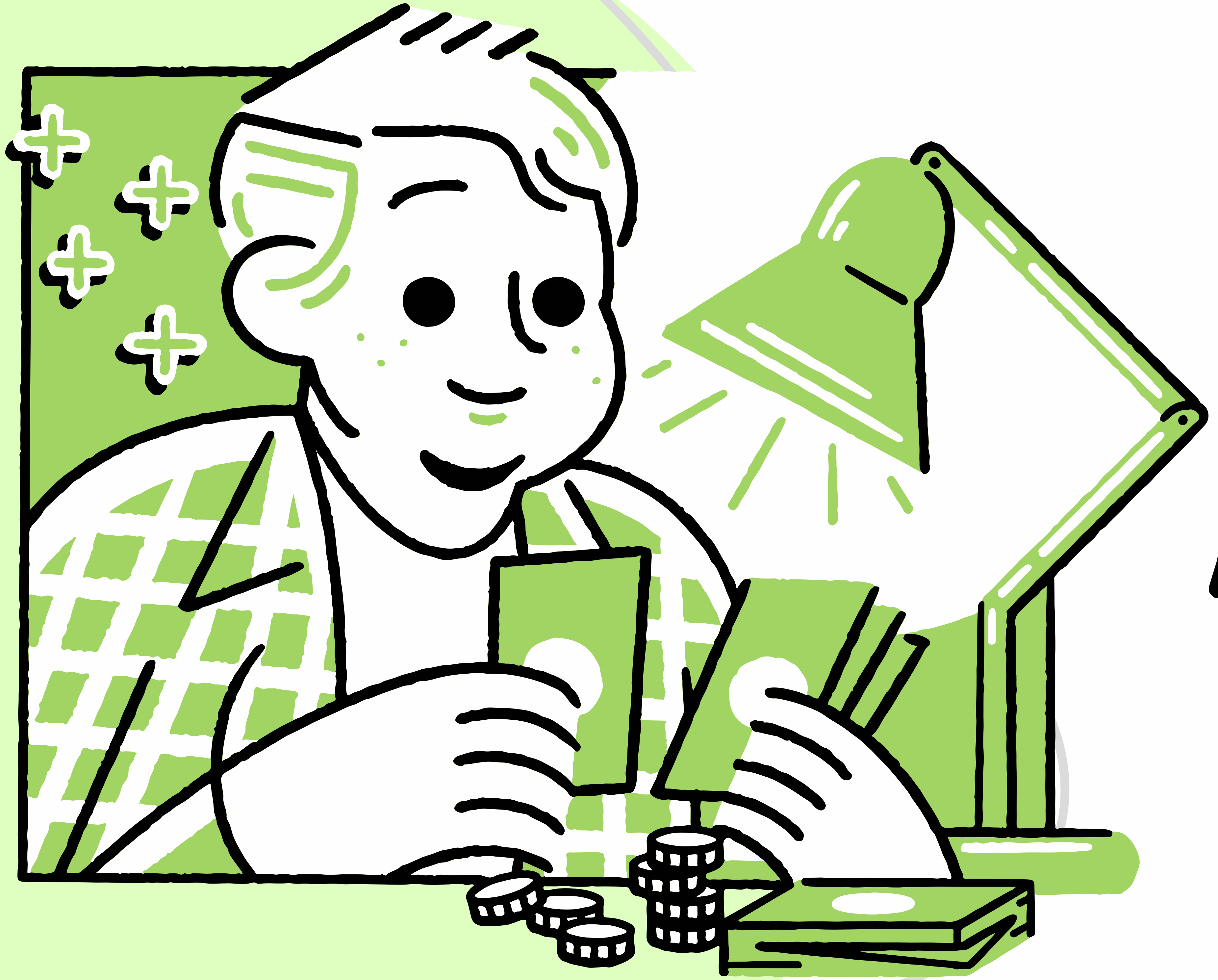
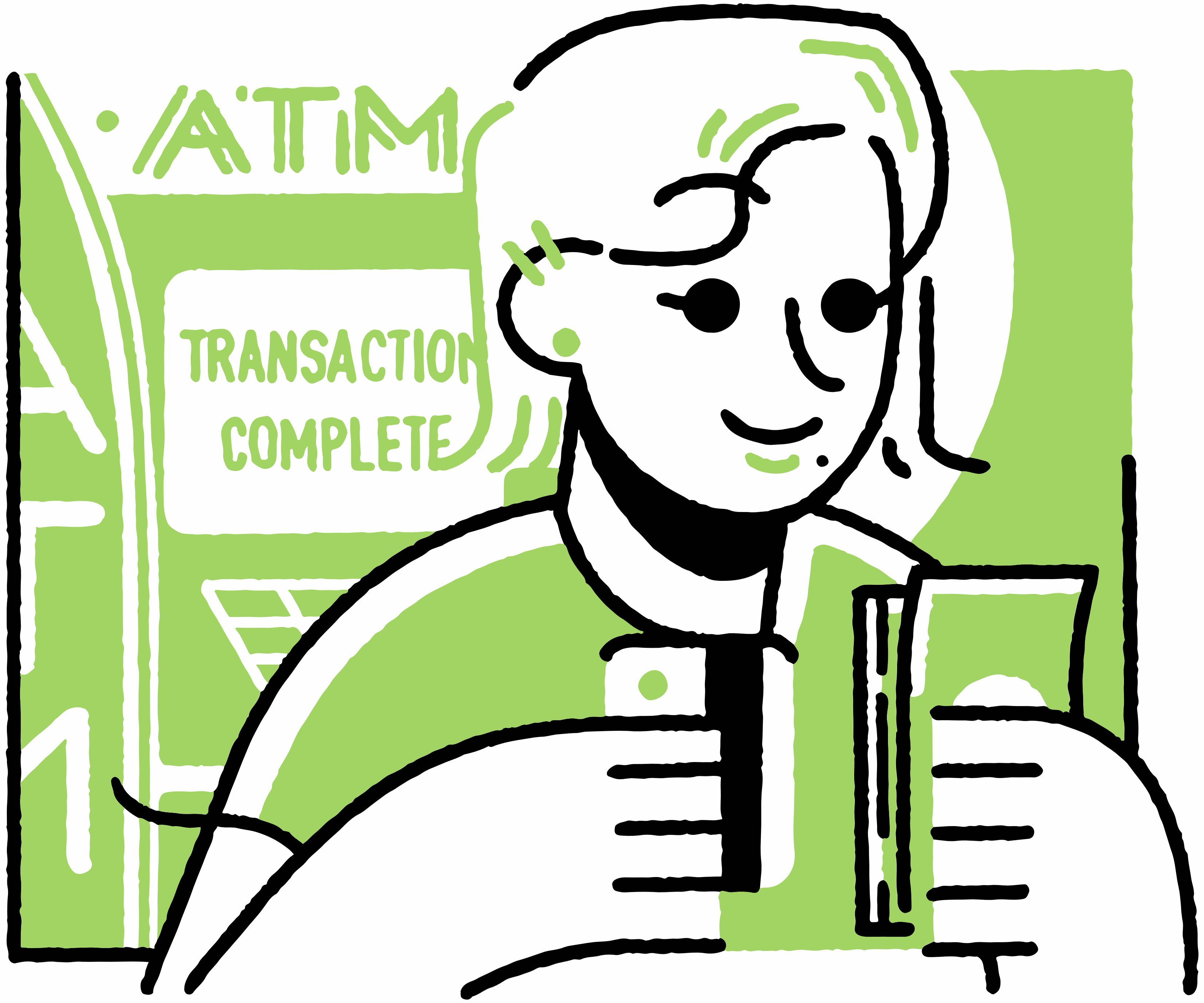


GROUP 3

CREDIT RISK ASSESSMENT

Group Members:
Adhiraj Singh
Ashmit Sharma
Divyanshu Gautam
Kanishk Tiwari



PROBLEM STATEMENT

In an era where lending is both essential for economic growth and fraught with risks, the challenge lies in discerning borrowers' creditworthiness accurately. Our mission is to harness the power of machine learning to create a predictive model that identifies the likelihood of loan default. By analyzing key borrower attributes and loan characteristics

RESEARCH PAPER-1

Bank credit risk analysis with k-nearest neighbor classifier: Case of Tunisian banks

Issue-1, Volume-14, 2015

- Research paper shows the application and usage of KNN classifier in assessing a person credit risk.
- Techniques implemented on 924 credit files of Tuisian bank with multi-dimensional data giving better results than simple one.
- The goal is to develop models capable of determining whether to approve or decline a credit application with minimum error rates, often by using Larger no. of k.
- Those attributes in the Dataset are chosen by observing linear regression models between many of them and, also to jus
- Selected Euclidean metric an adjusted version of the metric that contours for class membership. Also had a data-dependent metric.
- Dynamic updation in design set is a potential attraction to use KNN method.

ANALYSIS

Figure 2. ROC curve of three models

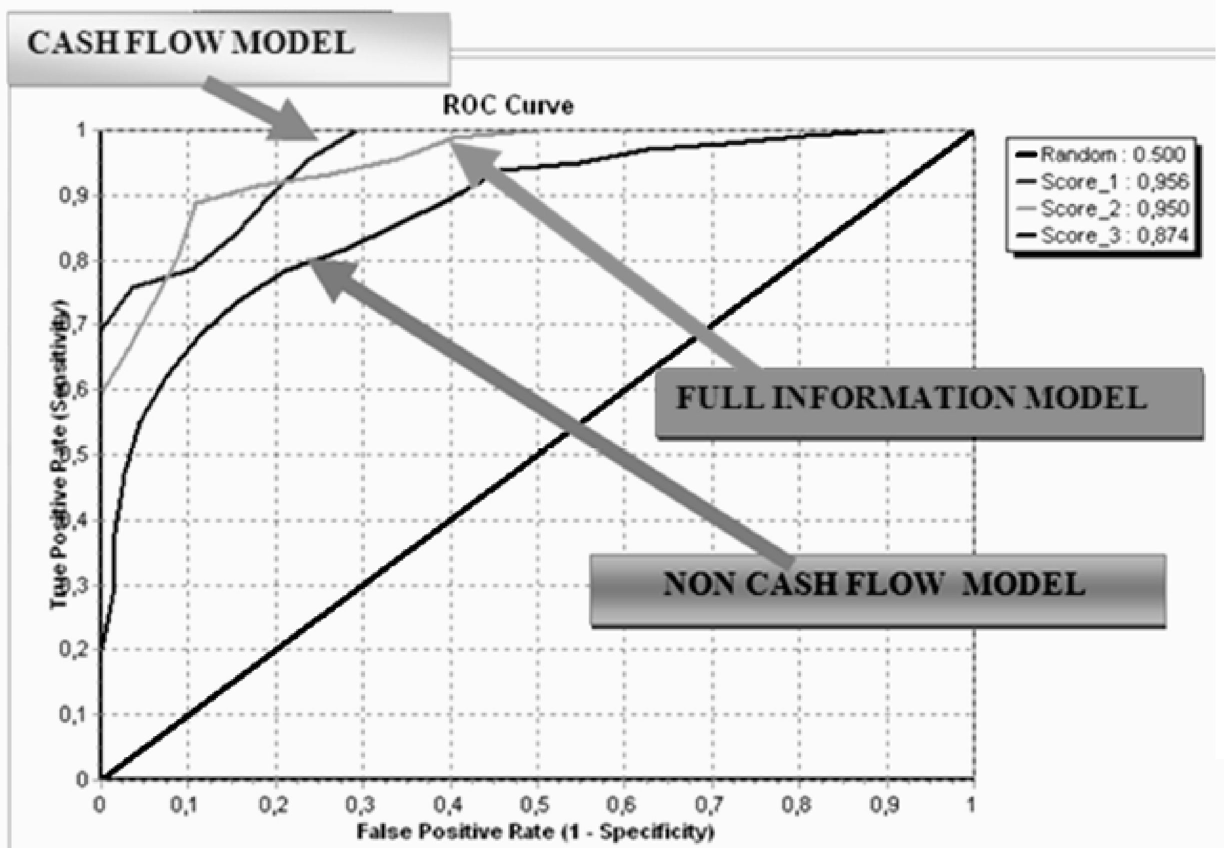


Table 5. Results for full information models (Appendix3 panels 1, 2, 3 and 4)

	K=2		K=3		K=4		K=5	
	Healthy	**Risky**	**Healthy**	**Risky**	**Healthy**	**Risky**	**Healthy**	**Risky**
Healthy companies	**393**	**65**	**406**	**52**	**381**	**77**	**383**	**75**
Risky companies	**69**	**397**	**69**	**397**	**99**	**367**	**113**	**353**
% Total Good and Bad Classification								
Good classification	**85.5%**		**86.90%**		**80.95%**		**79.65%**	
Bad classification	**14.50%**		**13.10%**		**19.05%**		**20.35%**	

Table 6. Criterion of the type I and II error

	ERROR	K=2	K=3	K=4	K=5
NON CASH FLOW MODEL	Type I	**21.83%**	**16.73%**	**27.51%**	**27.25%**
	Type II	**21.45%**	**20.52%**	**26.18%**	**27.72%**
CASH FLOW MODEL	Type I	**12.66%**	**12.01%**	**15.45%**	**19.74%**
	Type II	**13.75%**	**10.69%**	**15.50%**	**18.12%**
FULL INFORMATION MODEL	Type I	**14.8%**	**14.80%**	**21.24%**	**24.24%**
	Type II	**14.19%**	**11.35%**	**16.81%**	**16.37%**

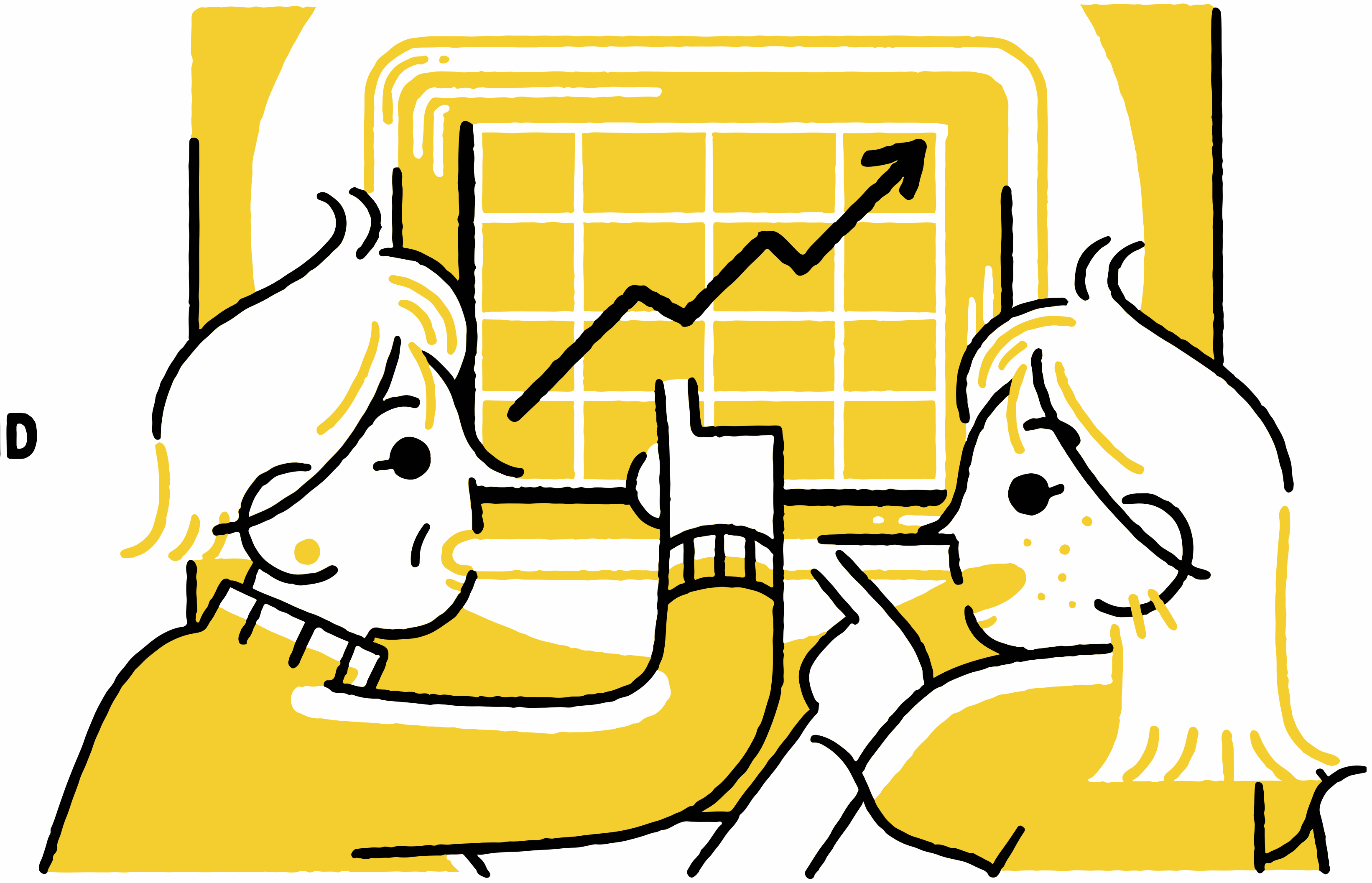
CONCLUSION FOR RESEARCH PAPER -1

- The research found that the best information set for credit risk assessment was related to accrual and cash flow, with a good classification rate achieved.
- Moreover, the Area Under Curve criterion was used to evaluate the performance of the model, with a reported AUC of 95.6% for the best model with cash flow information.

RESEARCH PAPER-2

CONFERENCE: THE ICCGANT 2020 CONFERENCE
PUBLISHED BY: JOURNAL OF PHYSICS

AUTHORS: N H PUTRI, M FATEKUROHMAN, AND I M TIRTA
FROM THE MATHEMATICS DEPARTMENT, FACULTY OF MATHEMATICS AND NATURAL SCIENCES, UNIVERSITY OF JEMBER, INDONESIA.



Credit Risk Analysis Using Support Vector Machines

- The paper highlights that SVM's ability to find the optimal hyperplane separating the classes with maximum margin makes it effective for credit risk classification compared to traditional statistical methods.
- SVM can handle high-dimensional, non-linear data through SVM kernels, which is useful for the multi-featured customer data involved in credit risk modeling.
- The paper tested SVM with four different kernel functions (linear, polynomial, RBF, sigmoid) and found that several of the SVM models achieved high accuracy rates around 0.95 in correctly classifying customers into good or bad credit classes using the bank's dataset.
- The SVM polynomial model also had a good balance of high sensitivity (0.9259), specificity (0.9579), precision (0.8621), and low false positive (0.0421) and false negative (0.0741) rates, which is important for minimizing bank losses from misclassifications.

Table 4. Model evaluation values fromprediction the testing data.

	ACC	TPR	TNR	PPV	FPR	FNR	F1-score
	0.9262	0.8889	0.9368	0.8000	0.0632	0.1111	0.8421
l	**0.9508**	**0.9259**	**0.9579**	**0.8621**	**0.0421**	**0.0741**	**0.8929**
	0.8934	0.9259	0.8842	0.6944	0.1158	0.0741	0.7937
	0.8361	0.8148	0.8421	0.5946	0.1579	0.1852	0.6875

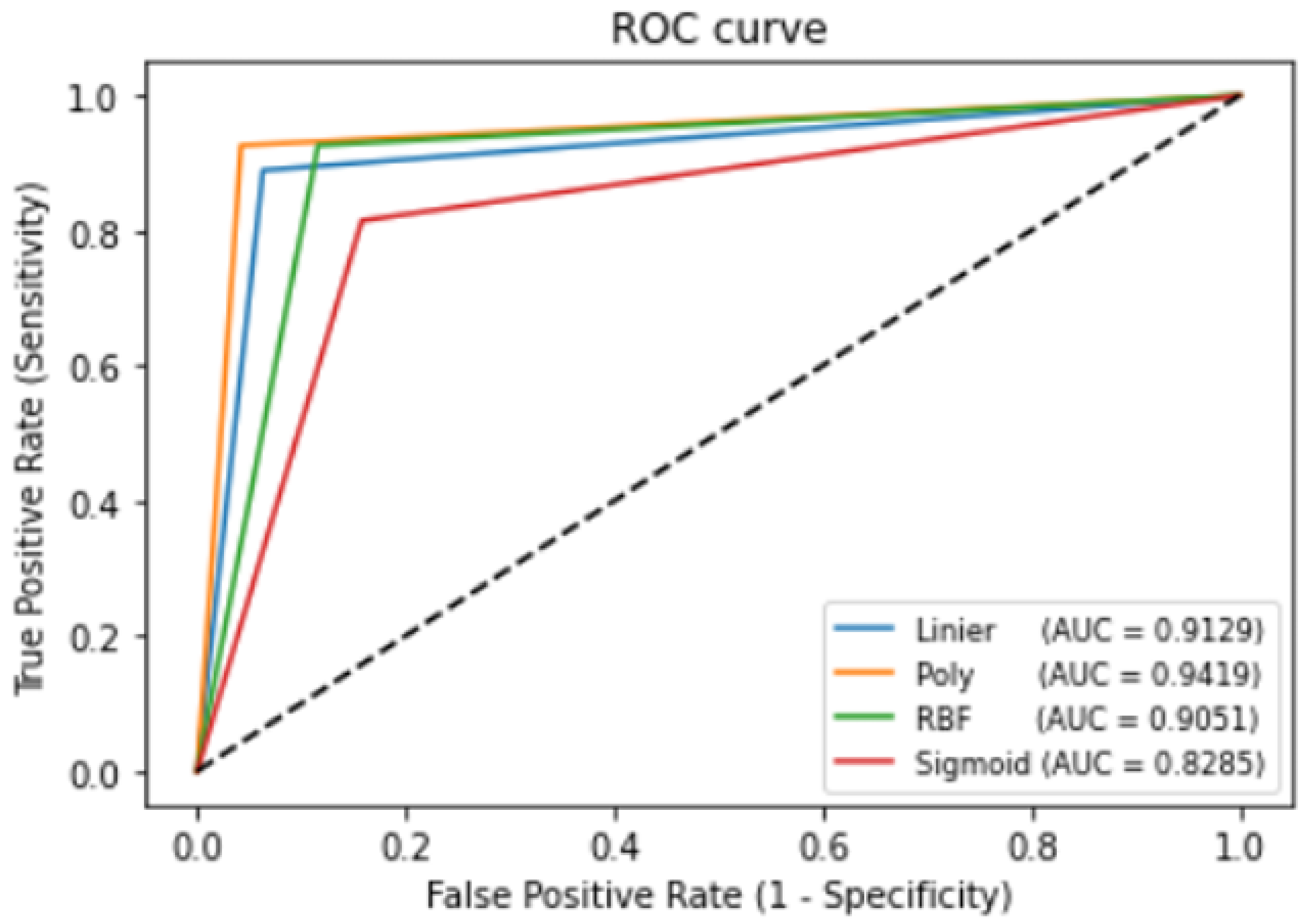


Figure 4. ROC Curves of SVM.

- Used dataset of 610 customer records from a bank from 2015-2018
- Independent variables: gender, loan amount, interest rate, loan term, job, income, collateral amount, loan history

Results and Conclusions

- SVM with polynomial kernel performed best:
  - Accuracy: 0.9508
  - AUC: 0.9419 (excellent classification)
  - High sensitivity (0.9259), specificity (0.9579), precision (0.8621)
  - Low FPR (0.421) and FNR (0.741)
- Polynomial SVM model can effectively classify credit applicants
- Can assist bank in accepting/rejecting applications to reduce bad credit risk

RESEARCH PAPER

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

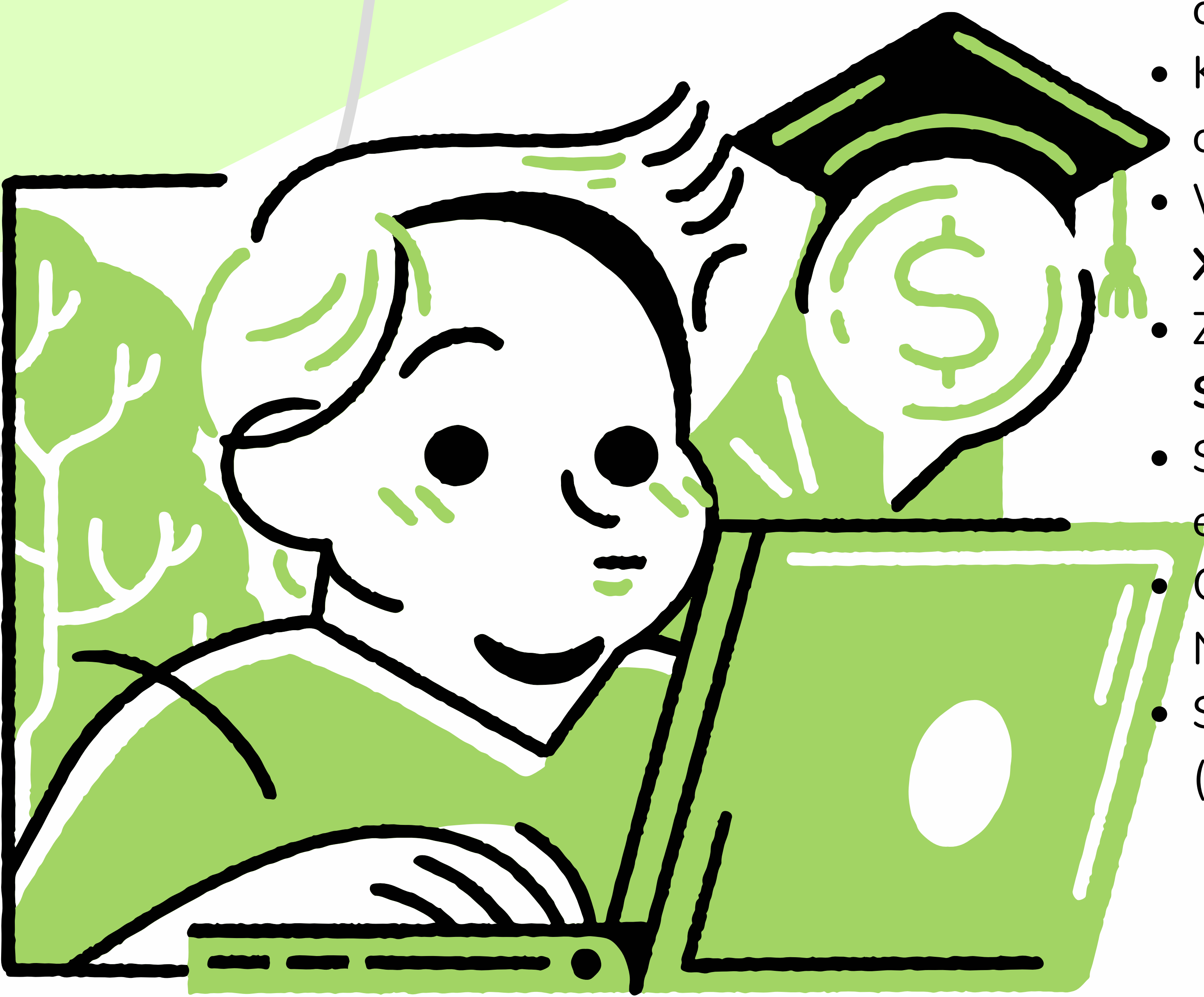
RESEARCH PAPER

- Research paper presents a comparison of various machine learning techniques for evaluating credit risk.
- Techniques implemented on the **German credit** dataset from the **UCI repository**.
- Dataset comprises 1000 rows and 21 columns.
- The objective is to train models to accept or reject credit profile based on attributes.
- Machine learning algorithms compared:
  - **Support Vector Network**
  - **Neural Network**
  - **Logistic Regression**
  - **Naive Bayes**
  - **Random Forest**
  - **Classification and Regression Trees (CART) algorithm**
- Results indicate that the Random Forest algorithm outperformed others in predicting credit risk with higher accuracy

RESEARCH PAPER

ATTRIBUTE NUMBER	DESCRIPTION	CLASS
1)	Creditability	Categorical
2)	Account Balance	Categorical
3)	Credit length (in months)	Numeric
4)	Status of payment	Categorical
5)	Purpose	Categorical
6)	Credit Amount	Numeric
7)	Savings in cost	Categorical
8)	Current employment period	Categorical
9)	Installment	Numeric
10)	Sex and Marital Status	Categorical
11)	Guarantors	Categorical
12)	Current address duration	Numerical
13)	Most precious resources	Categorical
14)	Lifespan	Numeric
15)	Simultaneous loans	Categorical
16)	Type of house	Categorical
17)	Amount of loans from this bank	Numeric
18)	Employment	Categorical
19)	Number of dependents	Categorical
20)	Telephone	Categorical
21)	Foreign Workers	Categorical

PREVIOUS WORK



- Li, Shiue, and Huang: Utilized **Support Vector Machine (SVM)** for credit risk assessment.
- Khandani, Kim, and Lo: Employed **Classification and Regression Trees (CART)** in their evaluation of customer loan risk.
- Devasena: Explored various supervised learning classification techniques, including **IBk classifier, Kstar classifier, and LWL classifier**.
- Gulsoy and Kulluk: Investigated Random Trees, **simple CART, PART, J48, Fuzzy**, and **NBTrees** for credit risk assessment.
- Huang, Liu, and Ren: Utilized the **Probabilistic Neural Network (PNN)** approach for credit risk evaluation.
- Khashman: Examined the **Emotional Neural Network (EmNN)** model for automatic credit rating.
- Wang, Yu, and Ji: Compared ensemble models such as **Random Forest, Naive Bayes, XGBoost, and RF-Bagging.**
- Zhong, Miao, Shen, and Feng: Explored **Artificial Neural Networks (ANN)** and **Support Vector Machines (SVM)** for credit risk assessment.
- Shukla and Nanda: Developed the "**parallel social spider algorithm**" for effective credit evaluation with mixed data types.
- Claderia, Brandao, Campos, and Pereira: Considered Logistic Regression, Neural Networks, Bayesian Networks, and Random Forests for credit risk assessment.
- Soui, Gasmi, Smiti, and Ghédira: Analyzed Multi-objective Evolutionary algorithms (SMOPSO, NSAG-II, MOEA/D, SPEA-2) for rule-based credit risk models.

ANALYSIS AND FUTURE SCOPE

- The paper compares different machine learning techniques for credit risk evaluation using the German credit dataset.
- Techniques include LR, BN, NN, SVN, RF, and CART algorithms.
- Testing was conducted on the German credit dataset with a large number of transactions.
- **Random Forest** methodology yielded **higher accuracy** in credit risk evaluation.
- Future work may explore the effectiveness of various deep learning techniques to further improve accuracy.

Table- II: Comparative outcome of different machine algorithm applied on German credit dataset

	ERR	ACC	REC	SP	PREC	F1-S
LR	0.25	0.75	0.91	0.34	0.76	0.83
NB	0.23	0.77	0.78	0.67	0.91	0.84
NN	0.26	0.74	0.80	0.59	0.82	0.81
SVN	0.29	0.76	0.81	0.50	0.80	0.83
RF	**0.22**	**0.78**	0.80	**0.67**	**0.91**	**0.85**
CAR T	0.23	0.77	**0.93**	0.39	0.78	0.84

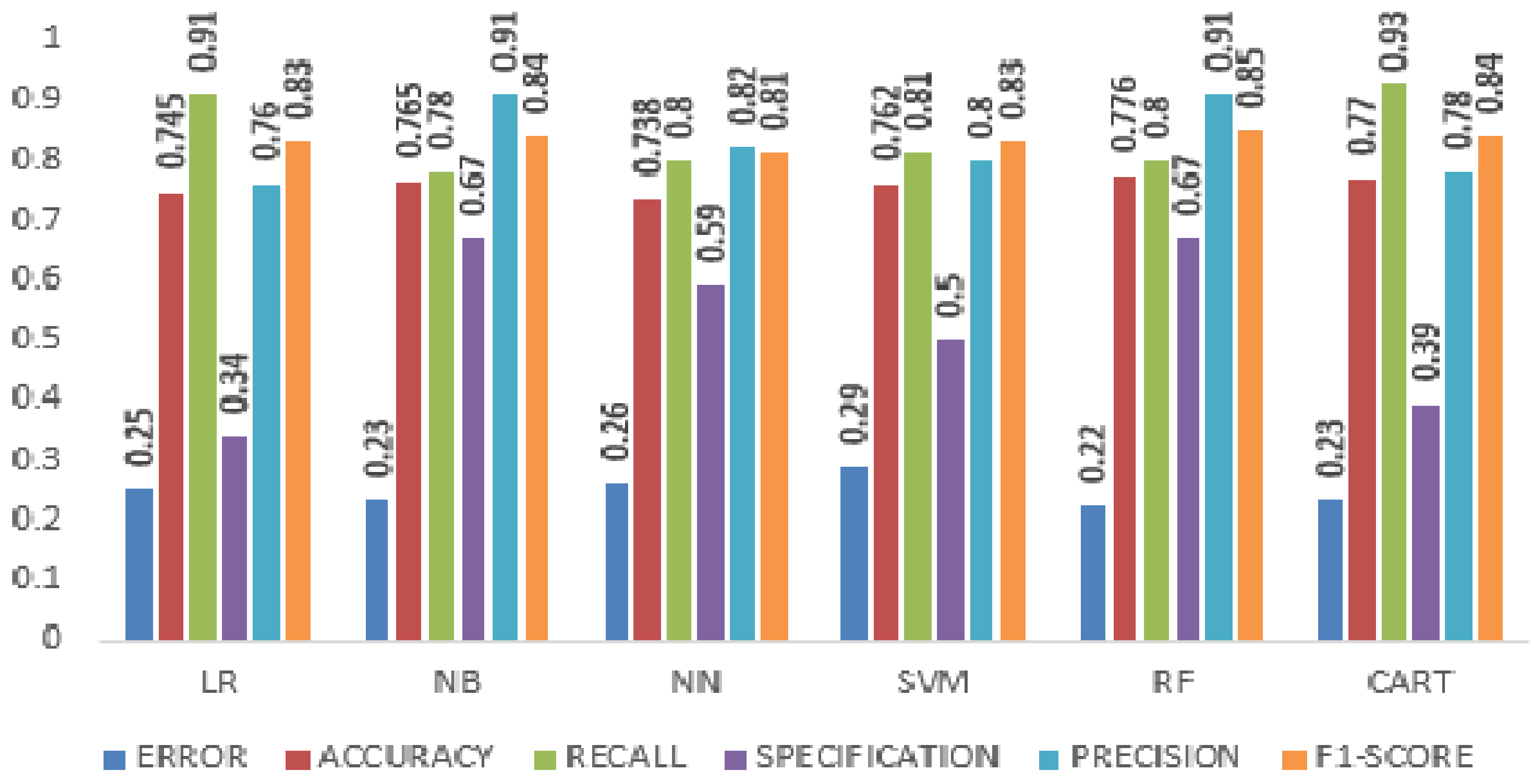


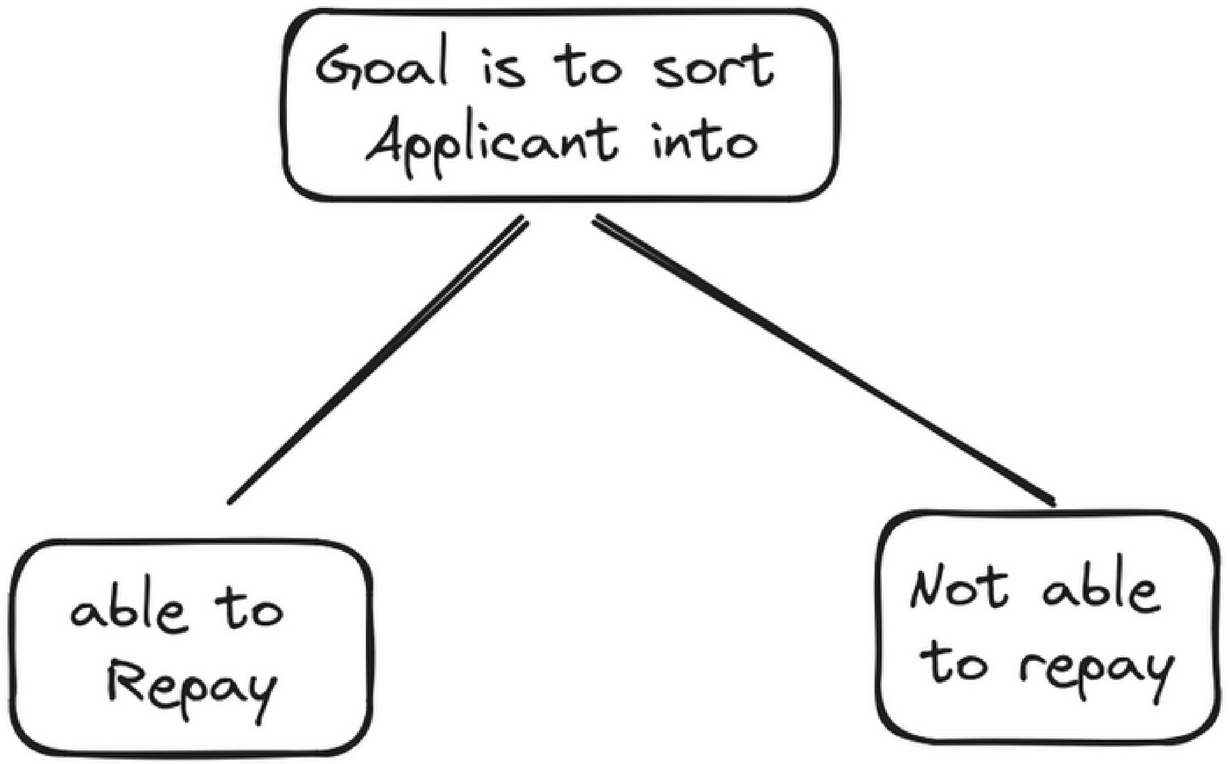
Fig. 3. Measures for different techniques

RESEARCH PAPER

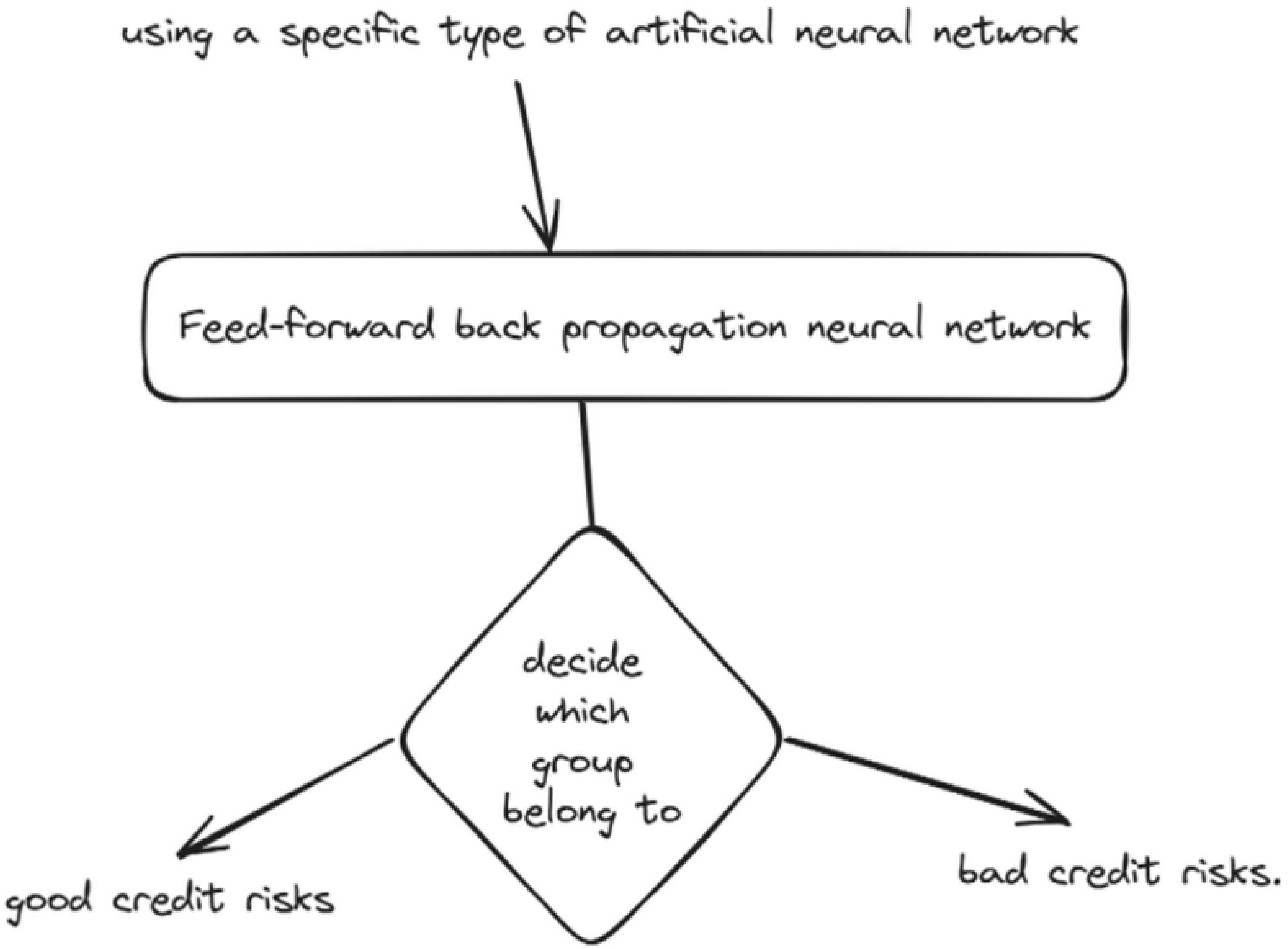
Global Engineers and Technologists Evaluating Credit Risk Using Artificial Neural Networks by Qeethara K. Al. Shayea, and Ghaleb A. El-Refae

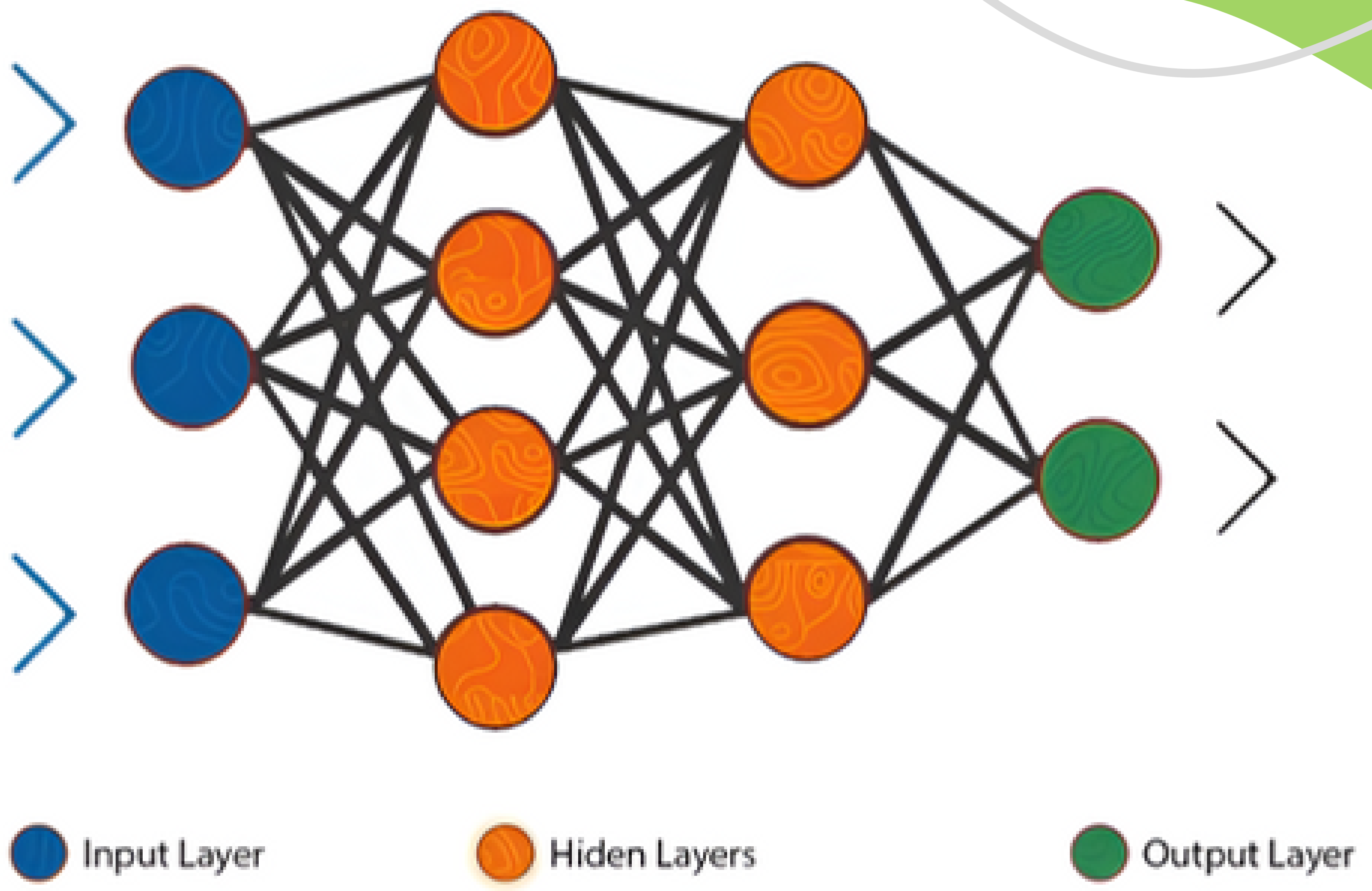
Publish on: September 2011

1 Objective



2 Method



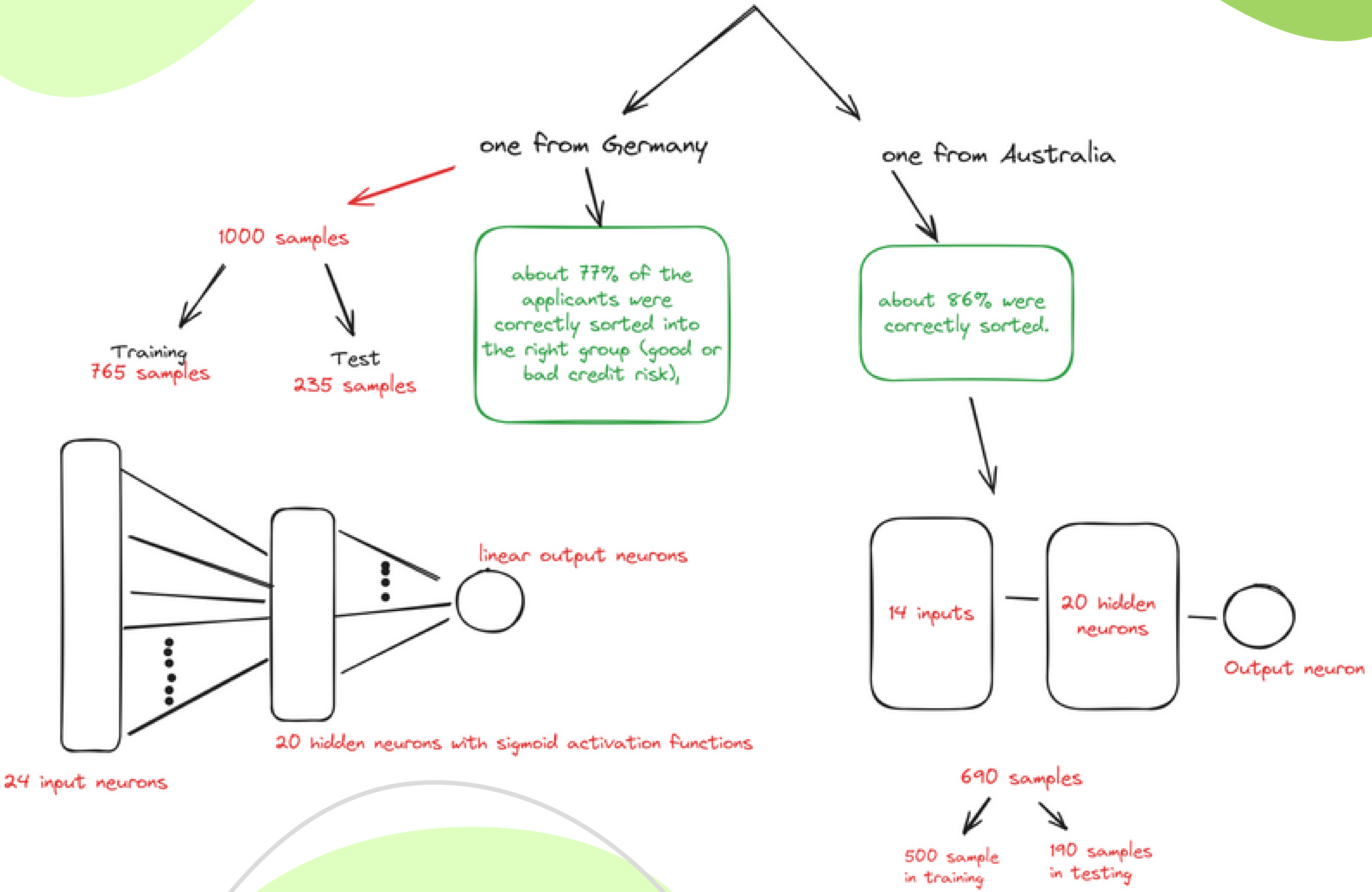


1\. In a feed-forward neural network, information flows in one direction only, from the input layer through one or more hidden layers to the output layer.

2\. There are no loops or feedback paths in this network.

4\. This setup allows the network to process data and make predictions based on the input without going back and forth.

3. Two Datasets



OUR APPROACH

- **Implementation of Artificial Neural Network (ANN):**
  - Utilizing ANN to enhance accuracy in data analysis.
  - Leveraging ANN's capability to learn complex patterns in data.
- **Feature Engineering:**
  - Conducting feature engineering to uncover hidden patterns within the dataset.
  - Identifying and extracting meaningful features to improve model performance.
- **Addressing Class Imbalance:**
  - Employing oversampling or undersampling techniques to handle imbalanced classes.
  - Ensuring fair representation of all classes to avoid bias in model training.
- **Comparison of Methods:**
  - Analyzing the effectiveness of conventional methods versus neural networks.
  - Drawing insights from relevant research papers to inform our approach.
- **Enhancing Accuracy:**
  - Integrating findings from research papers to optimize model performance.
  - Striving for higher accuracy through a combination of methods and techniques.