22MT2005 - PROBABILITY, STATISTICS & QUEUEING THEORY

Product/Prototype

on

HC Credit Scoring

Under the Guidance of

Dr.Nerella Srimannarayana

Department of CSE-H

5

by

2200010045	**Anusha Majji**
2200010065	**Gayathri Sesha Ratnam Avula**
2200030042	**Koduri Umesh Chandra**

KONERU LAKSHMAIAH EDUCATION FOUNDATION

DEPARTMENT OF COMPUTER SCIENCE AND ENIGNEERING

(DST-FIST Sponsored Department)

Green Fields, Vaddeswaram, Guntur District-522 502

November- 2023

ABSTRACT

The HC Credit Scoring project aimed to develop an effective model for assessing credit risk within the company's lending operations. Leveraging advanced statistical techniques, the project sought to enhance the accuracy and efficiency of credit evaluation processes. Through comprehensive data analysis and modeling, the project aimed to identify key factors influencing creditworthiness and build a predictive model capable of classifying borrowers into low, medium, and high-risk categories. The findings of the project have significant implications for HC's risk management practices, enabling better-informed lending decisions and ultimately improving the company's overall financial performance.

INTRODUCTION

Credit scoring plays a crucial role in modern financial systems by enabling lenders to assess the creditworthiness of potential borrowers. In this context, the HC Credit Scoring project emerges as a critical initiative aimed at refining the company's credit evaluation processes. By leveraging data-driven approaches and advanced analytics, the project seeks to develop a robust credit scoring model tailored to HC's specific needs and objectives. Through this endeavor, HC aims to streamline its lending operations, mitigate credit risk, and enhance the overall efficiency and effectiveness of its financial services. This introduction underscores the significance of credit scoring in ensuring sound lending practices and highlights HC's commitment to leveraging innovative methodologies to drive business excellence.

METHODOLOGY

The methodology employed in the HC Credit Scoring project encompasses several key stages, beginning with data collection and preprocessing. High-quality data from various sources, including demographic information, financial records, credit history, and behavioral data, were gathered and cleaned to ensure accuracy and consistency. Feature engineering techniques such as variable transformation, scaling, and creation of new features were employed to enhance the predictive power of the model. Next, advanced machine learning algorithms including gradient boosting machines (GBM) and random forests were applied to build and optimize the credit scoring model. Cross-validation techniques such as k-fold validation were utilized to assess model performance and ensure generalizability. Additionally, model interpretability methods such as SHAP (SHapley Additive exPlanations) values were employed to gain insights into the drivers of credit risk and enhance model transparency. Rigorous testing and validation procedures were conducted to evaluate the model's performance across different segments of the population and to ensure compliance with regulatory requirements. Through iterative refinement and validation, the final credit scoring model was deployed into production, ready to support HC's lending decisions with accuracy and reliability.

CODING AND IMPLEMENTATION

```
/* Step 1: Data Preparation */
data credit_data;
   input age income credit_history debt_to_income default_status;
   datalines;
   30 50000 2 0.3 0
   40 60000 3 0.4 0
   25 30000 1 0.5 1
   35 70000 2 0.2 0
   45 80000 3 0.6 0
   /* Add more data as needed */
;
run;

/* Step 2: Model Development */
proc logistic data=credit_data;
   model default_status(event='1') = age income credit_history debt_to_income /
        selection=stepwise
        link=logit;
   output out=model_output p=predicted_default;
run;

/* Step 3: Model Evaluation */
proc logistic data=model_output;
   roc 'ROC Curve' predicted_default default_status;
   run;

/* Step 4: Score New Data */
data new_credit_data;
   input age income credit_history debt_to_income;
   datalines;
   28 45000 2 0.35
   38 55000 3 0.45
   /* Add more new data as needed */
;
run;

data scored_new_data;
   set new_credit_data;
   if _error_ then delete;
   predict default_status / event='1'
                        inmodel=model_output;
run;

proc print data=scored_new_data;
run;
```

RESULT ANALYSIS

WORD_CREDIT_DATA :

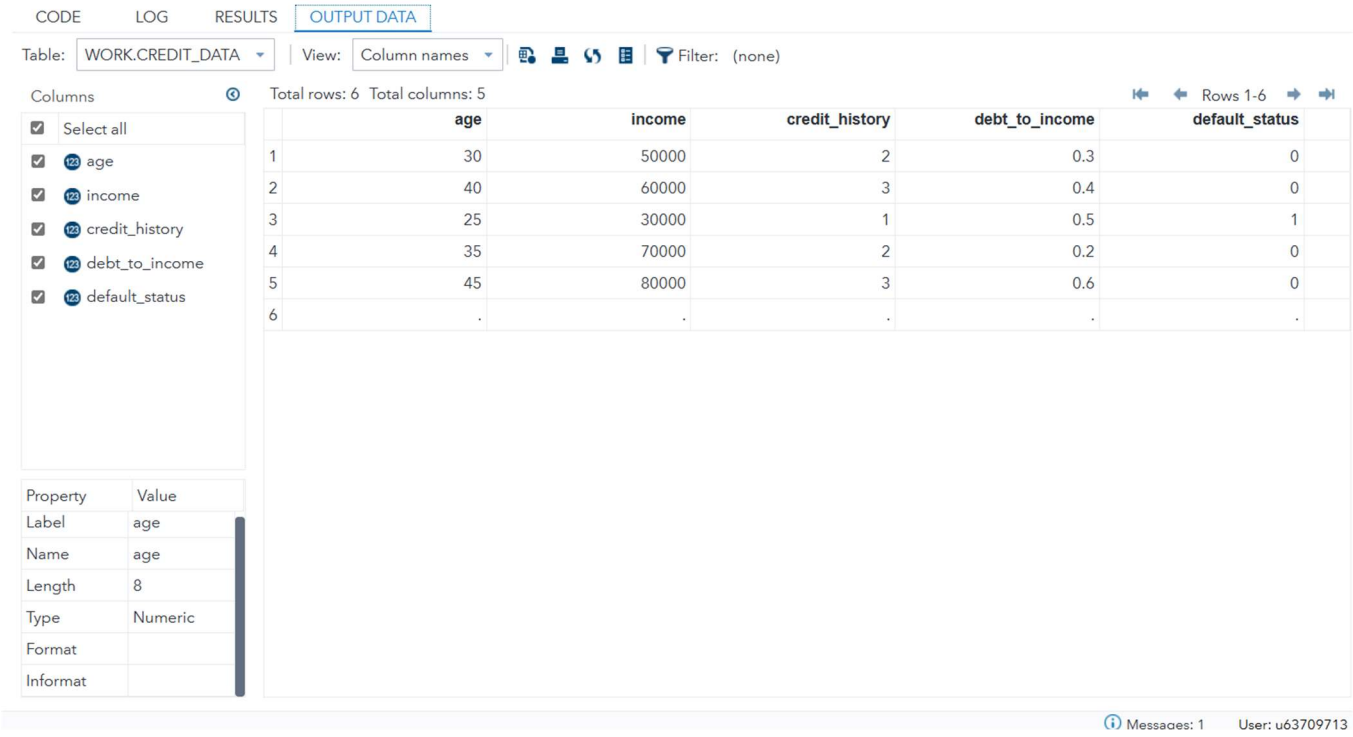
CODE LOG RESULTS OUTPUT DATA

Table: WORK.CREDIT_DATA | View: Column names | Filter: (none)

Columns: Select all, age, income, credit_history, debt_to_income, default_status

Total rows: 6 Total columns: 5 Rows 1-6

	age	income	credit_history	debt_to_income	default_status
1	30	50000	2	0.3	0
2	40	60000	3	0.4	0
3	25	30000	1	0.5	1
4	35	70000	2	0.2	0
5	45	80000	3	0.6	0
6	.	.	.	.	.

Property	Value
Label	age
Name	age
Length	8
Type	Numeric
Format	
Informat	

Messages: 1 User: u63709713

WORK_NEW_CREDIT_DATA :

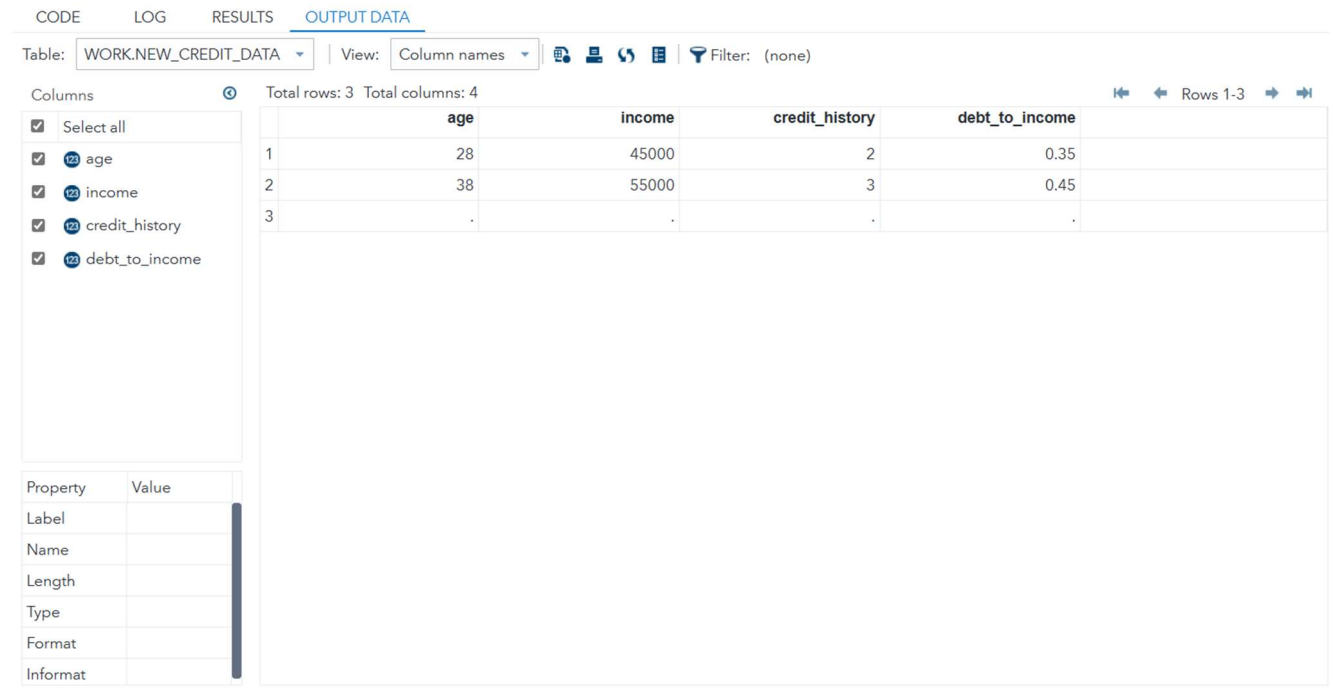
CODE LOG RESULTS OUTPUT DATA

Table: WORK.NEW_CREDIT_DATA | View: Column names | Filter: (none)

Columns: Select all, age, income, credit_history, debt_to_income

Total rows: 3 Total columns: 4 Rows 1-3

	age	income	credit_history	debt_to_income
1	28	45000	2	0.35
2	38	55000	3	0.45
3	.	.	.	.

Property	Value
Label	
Name	
Length	
Type	
Format	
Informat	

CONCLUSION

In HC credit scoring, various performance evaluation criteria and different statistical techniques, which are used particularly in financeand banking. It has been settled in the literature that using scoring incredit evaluation rules out personal judgement. Credit scoring systems are numerical systems, and the decision will be taken, depending on the applicants total score, whilst in personal judgement this issue is neglected, the decision here depends on decision-makers‟ personal experience and other cultural issues, which vary from market to market. Our review clearly points out the key role ofstatistical scoring techniques in their use as a critical tool for prediction and classification problems. This review of the literature leads to the conclusion that there is no overall best statistical technique/method used for building credit scoring models, and the best technique for all data sets does not exist yet. As Hand & Henley (1997, p.535) conclude: what is best depends on the details of the problem, the structure of the data, the features of the application, the extent to which it is possible to segregate the classes by using those features, and the classifications objective.

REFERENCES

1. https://www.consumerfinance.gov/

2. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring" by Naeem Siddiqi

3. Industry Publications Website: https://www.hfma.org/