National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Pranavi Pusapati
Student ID: x20155301

School of Computing
National College of Ireland

Supervisor: **Jorge Basilio**

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Pranavi Pusapati

Student ID: x20155301

Programme: Data Analytics **Year:** 2021

Module: Configuration Manual

Supervisor: Jorge Basilio

Submission Due Date: 16/12/2021

Project Title: Loan Under writing prediction using Deep learning techniques

Word Count: 1160 **Page Count** 9

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Pranavi Pusapati

Date: 16/12/2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

PRANAVI PUSAPATI
x20155301

1. Introduction

The following configuration manual illustrates the requirements for implementing the system which was designed for predicting the loan under writing by using the Deep Learning models and dropout function. Further, the manual will thoroughly explain the software and hardware requirements that were used for the successful implementation of the project.

2. System Configuration

Following are the hardware and software configuration which were used for the implementation of this Project.
The hardware configurations used for implementation are as follows:

2.1. Hardware Requirement

Table 1: Hardware Requirements

Hardware	Configurations
System	Lenovo Z580 Idea pad
Operating System	Windows 10 (64 Bits) Pro
RAM	4 GB
Hard Disk	1 TB
Graphics Card	NVIDIA RTX 2060 (6 GB)
Processor	Intel Core i5-3230M

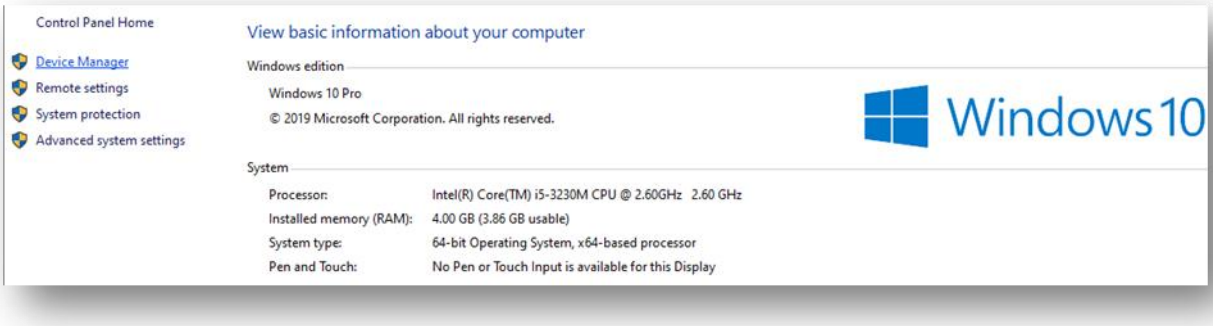


Figure 1: Operating System Configurations

2.2. Software Requirement

Table 2: Software Requirements

Software	Version
Python	3.8 (64 Bits)
Google Colab Community	2021.2 (64 Bits)

Figure 2: Google Collaboratory with Python

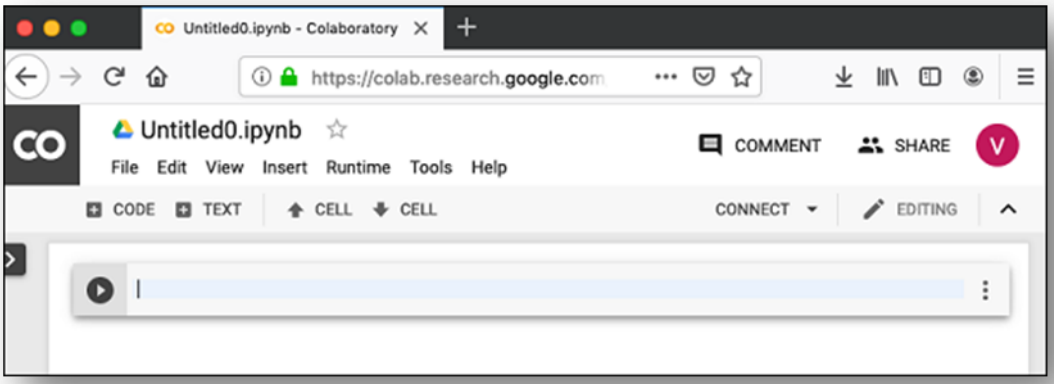


Figure 3: Python code files in Google Colab

3. Project Implementation

3.1. Data Collection

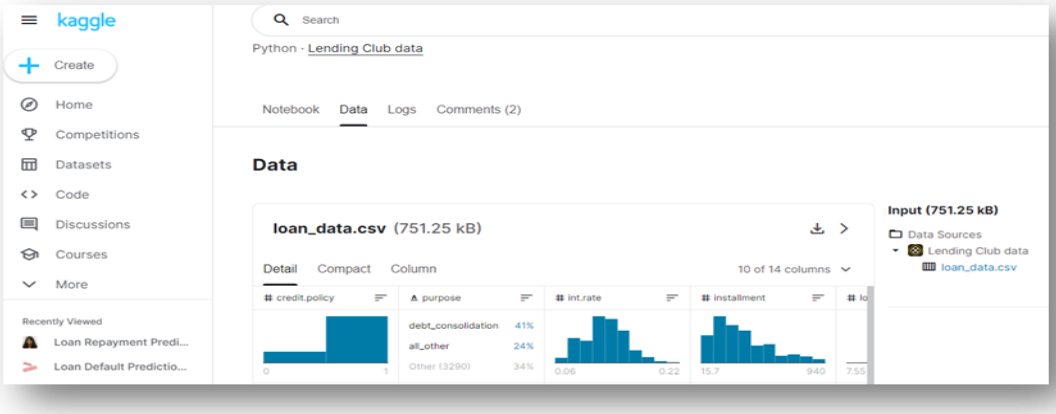


Figure 4: Data Collection from the source

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9578 entries, 0 to 9577
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   credit.policy      9578 non-null   int64
 1   purpose            9578 non-null   object
 2   int.rate           9578 non-null   float64
 3   installment        9578 non-null   float64
 4   log.annual.inc     9578 non-null   float64
 5   dti                9578 non-null   float64
 6   fico               9578 non-null   int64
 7   days.with.cr.line  9578 non-null   float64
 8   revol.bal          9578 non-null   int64
 9   revol.util         9578 non-null   float64
 10  inq.last.6mths     9578 non-null   int64
 11  delinq.2yrs        9578 non-null   int64
 12  pub.rec            9578 non-null   int64
 13  not.fully.paid     9578 non-null   int64
dtypes: float64(6), int64(7), object(1)
memory usage: 1.0+ MB
```

Figure 5: Data description analysis

3.2. Data Description

credit_policy: 1 if the customer meets the credit underwriting criteria and 0 otherwise.
purpose: The purpose of the loan such as: credit_card, debt_consolidation, etc.
int_rate: The interest rate of the loan (proportion).
installment: The monthly installments ($) owed by the borrower if the loan is funded.
log_annual_inc: The natural log of the annual income of the borrower.
dti: The debt-to-income ratio of the borrower.
fico: The FICO credit score of the borrower.
days_with_cr_line: The number of days the borrower has had a credit line.
revol_bal: The borrower's revolving balance.
revol_util: The borrower's revolving line utilization rate.
inq_last_6mths: The borrower's number of inquiries by creditors in the last 6 months.
delinq_2yrs: The number of times the borrower had been 30+ days past due on a payment in the past 2 years.
pub_rec: The borrower's number of derogatory public records.

not_fully_paid: indicates whether the loan was not paid back in full (the borrower either defaulted or the borrower was deemed unlikely to pay it back).

3.3. Data Pre-processing

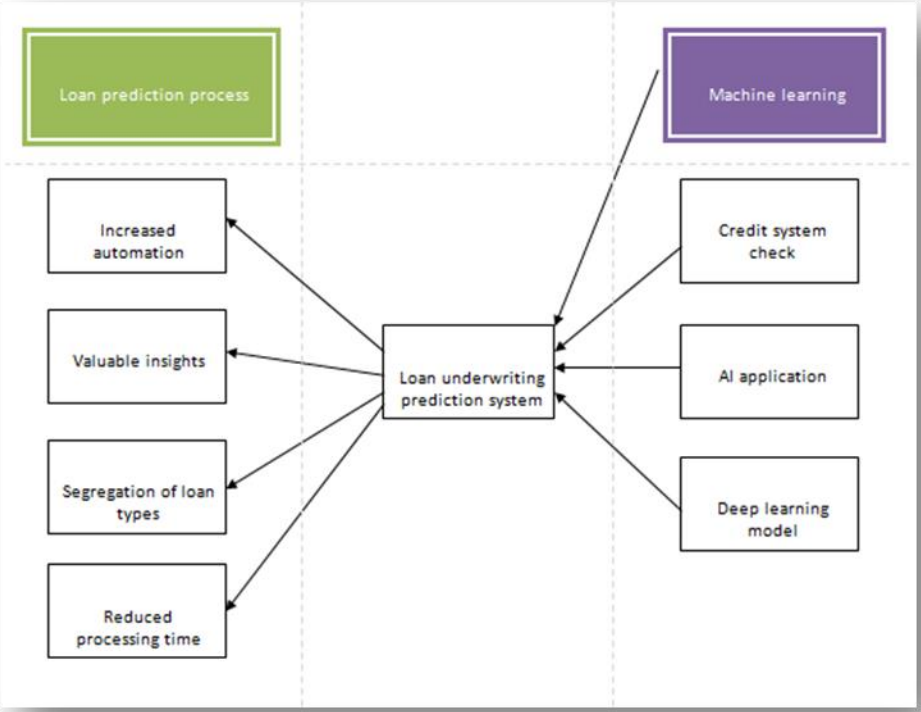


Figure 6: Conceptual framework

S.R.	Feature selection based on customer behavior	Description of features
1	loan_policy	1 If the customer meets the loan underwriting criteria, and 0 otherwise
2	type_of_purpose	This refers to the purpose of the loan
3	int_rate	The interest rate of the loan
4	days_	The no. of days to loan sanctioned
5	inq_	The loan holder inquiries
6	no_of_installment	The monthly installments owed by loan holder if the loan is sanctioned
7	loan_fully_paid_or_not	This indicates whether the loan was fully paid *r* or not

Figure 7: Feature selection and description

4. Model Building

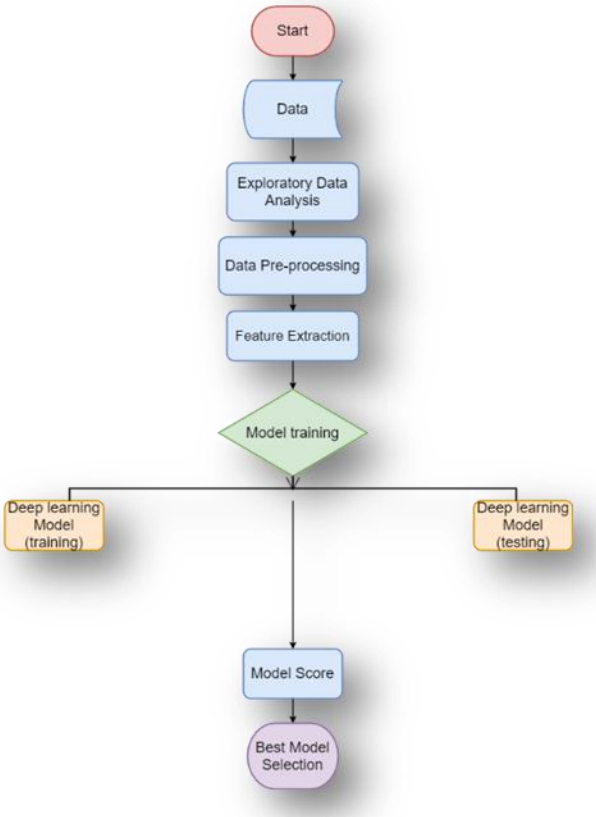
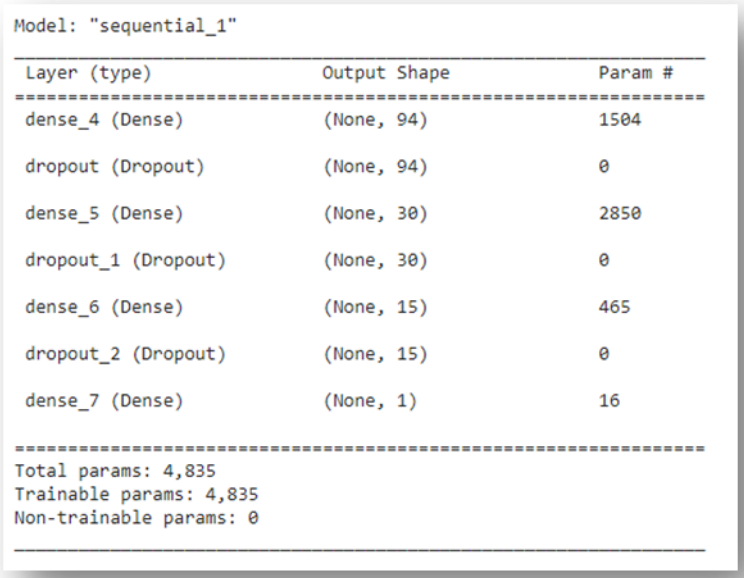


Figure 8: Methodology Diagram

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 94)                1504

 dense_1 (Dense)             (None, 30)                2850

 dense_2 (Dense)             (None, 15)                465

 dense_3 (Dense)             (None, 1)                 16

=================================================================
Total params: 4,835
Trainable params: 4,835
Non-trainable params: 0
_________________________________________________________________
```

Figure 9: Model summary for deep learning model without dropout function



```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_4 (Dense)             (None, 94)                1504

 dropout (Dropout)           (None, 94)                0

 dense_5 (Dense)             (None, 30)                2850

 dropout_1 (Dropout)         (None, 30)                0

 dense_6 (Dense)             (None, 15)                465

 dropout_2 (Dropout)         (None, 15)                0

 dense_7 (Dense)             (None, 1)                 16

=================================================================
Total params: 4,835
Trainable params: 4,835
Non-trainable params: 0
_________________________________________________________________
```

Figure 10: Model summary for deep learning model with dropout function

References

[1] Brahma, A., Goldberg, D. M., Zaman, N., & Aloiso, M. (2021). Automated mortgage origination delay detection from textual conversations. *Decision Support Systems*, *140*, 113433.

[2] Breeden, J. L., & Leonova, E. (2021). Creating Unbiased Machine Learning Models by Design. *Journal of Risk and Financial Management*, *14*(11), 565.

[3] Bwire, A. C., Tenai, J. K., & Odunga, R. M. (2021). Loan Growth, Loan Deposit Ratio and Prediction of Bank Fragility in Kenya Using Generalised Linear Model. *African Journal of Education, Science and Technology*, *6*(3), 103-113.

[4] Ekin, T. (2020). Discussion of "Machine learning applications in nonlife insurance". *Applied Stochastic Models in Business and Industry*, *36*(4), 541-544.

[5] Gikay, A. A. (2021). The American Way—Until Machine Learning Algorithm Beats the Law?. *Case Western Reserve Journal of Law, Technology & the Internet, Volule*, *12*, 1-56.

[6] Goldberg, D. M., Zaman, N., Brahma, A., & Aloiso, M. (2021). Are mortgage loan closing delay risks predictable? A predictive analysis using text mining on discussion threads. *Journal of the Association for Information Science and Technology*.

[7] Knight, E. (2019). AI and Machine Learning-Based Credit Underwriting and Adverse Action under the ECOA. *Bus. & Fin. L. Rev.*, *3*, 236.

[8] Kumar, A., Shanthi, D., & Bhattacharya, P. (2021, August). Credit Score Prediction System using Deep Learning and K-Means Algorithms. In *Journal of Physics: Conference Series* (Vol. 1998, No. 1, p. 012027). IOP Publishing.

[9] Li, J. P., Mirza, N., Rahat, B., & Xiong, D. (2020). Machine learning and credit ratings prediction in the age of fourth industrial revolution. *Technological Forecasting and Social Change*, *161*, 120309.

[10] Munoz, J., Rezaei, A. A., Jalili, M., & Tafakori, L. (2021). Deep learning based bi-level approach for proactive loan prospecting. *Expert Systems with Applications*, *185*, 115607.

[11] Pallathadka, H., Ramirez-Asis, E. H., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J. M., & Naved, M. (2021). Applications of artificial intelligence in business management, e-commerce and finance. *Materials Today: Proceedings*.

[12] Perkins, D. W. (2018). Marketplace lending: Fintech in consumer and small-business lending. *Congressional Research Service*, *12*(7), 1-30.

[13] Pillai, S. G., Woodbury, J., Dikshit, N., Leider, A., & Tappert, C. C. (2019, October). Machine Learning Analysis of Mortgage Credit Risk. In *Proceedings of the Future Technologies Conference* (pp. 107-123). Springer, Cham.

[14] Rezina, S. (2020). Non-Performing Loan in Bangladesh: A Comparative Study on the Islamic Banks and Conventional Banks. *Indian Journal of Finance and Banking*, *4*(1).

[15] Sabharwal, C. L. (2018). The rise of machine learning and robo-advisors in banking. *IDRBT Journal of Banking Technology*, 28.

[16] Sachan, S., Yang, J. B., Xu, D. L., Benavides, D. E., & Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. *Expert Systems with Applications*, *144*, 113100.

[17] Shrivastava, A. (2018). Usage of Machine Learning In Business Industries and Its Significant Impact. *International Journal of Scientific Research in Science and Technology*, *4*(8).

[18] Tappert, C. C. (2019, October). Machine Learning Analysis of Mortgage Credit Risk. In *Proceedings of the Future Technologies Conference (FTC) 2019: Volume 1* (Vol. 1069, p. 107). Springer Nature.