IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# APPLICATION OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING TO PREDICT GROUND RENT FEES IN ZAMBIA

**Author(s) DEAN J. NALUMPA, MSc Computer Science (Candidate), DMI St Eugene University, Lusaka. Zambia**

**DR K. M. ABUBAKKAR SITHIK, Research Coordinator, St John Baptist University, Blantyre. Malawi.**

**DR TAWARISH Assistant Professor, Department of information and system management, The NEW College Chennai. INDIA.**

## Abstract

The administration of land in Zambia largely refers to the process of finding available land, surveying it and allocating to citizens or investors that will develop the land. Citizen's allocated land are expected to develop the land in eighteen months (18) and are required to pay annual statutory fees. Among them ground rent fees and this is according to the Ministry of Justice, Land Act of 1975. The ground rent fees are generated by the Land's Information System annually. Hardcopy bills are printed on demand and in some cases sent as bulk SMS's to property owners reminding them of outstanding ground rent fees. In a country like Zambia were resources are low and with properties well above one million and thirty-five thousand (1,035,000), it is very difficult to prioritize as it is not easy to identify which property categories should be targeted. The purpose of this study is to examine the possibility of applying Artificial Intelligence (AI) Supervised Machine Learning (ML) in the prediction of ground rent fees in a particular category or categories (i.e Commercial, Industrial, Residential, Agriculture..etc) for future planning purposes. Land properties from Lusaka, Southern, Copperbelt, Northern and North-Western Provinces will be used in the study as sample data. The predicted results will help identify applicable ground rent fees from the various properties in different categories and this will help prioritize resources and concentrate on categories of properties with future ground rent fees that are expected to be high. This will also help minimize resource wastage as only selected properties with high ground rent fees are targeted.

# 1. Introduction

Ground rent administration is the sole responsibility of the Ministry of Lands and Natural Resources (MLNR) in Zambia which has been tasked with the responsibility of land administration under the Land Act of 1975. The MLNR has the responsibility of identifying available land, surveying it and allocate to the citizens of the republic of Zambia or investors that are willing to develop the land and create employment. The ministry has three core functions which perform the following: Land Leasing; Surveying; and Issuance of Title Deeds. The Land Leasing functionality is responsible the allocation, leasing, inspections and generation and distribution of ground rent bills. Ground rent bills are a statutory fee that each property pays on an annual basis. This paper will therefore specifically focus on the ground rent administration component. There are different land uses defined in the Lands Act of 1975. The Land uses fall in the following categories: Low Residential; Medium Residential; High Residential; Commercial; Industrial; and Agricultural. Ground rent fee bills are generated for each category of land. The ground rent fee bills must be printed on demand by property owners and in some cases sent to owners via bulk. The process is largely affected by poor collection of revenue as the ministry is not able to identify and target properties in a given category with seemingly high ground rent fees. With the advent of AI technologies, it is imperative that AI Machine Learning Models are used to predict ground rent fees and help stop resource wastage, minimize user intervention and prioritize usage of resources. In order to test the applicability of AI in managing ground rent fees, a machine learning model will created, trained and tested. We will download Anaconda Navigator, an open source software that enables the software developer/Data Scientist to build Jupyter Notebook Machine Learning Models. But first before we can begin the building process it is important that the subject matter in this case AI is understood clearly.

## 1.1 What is AI?

Artificial Intelligence or sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals. Some of the activities that it is designed to do is speech recognition, learning, planning and problem solving. Artificial intelligence (AI) is a branch of computer science. It involves developing computer programs to complete tasks which would otherwise require human intelligence. AI algorithms can tackle learning, perception, problem-solving, language-understanding and/or logical reasoning. AI is used in many ways within the modern world, from personal assistants to self-driving car. Artificial intelligence (AI) is evolving rapidly. While science fiction every so often portraits AI as robots closely as possible to humans (Ziyad Mohammed, 2018/2019).

## 1.2 Types of AI

### 1.2.1 Type 1

Artificial intelligence today is accurately known as narrow AI or weak AI, it is non-sentient machine intelligence, typically designed to perform a narrow task e.g. only facial and speech recognition or only internet searches or only driving a car. However, the long-term goal of many researchers is to create an Artificial General Intelligence (AGI) or strong AI which is a machine with the ability to apply intelligence to any problem, rather than just one specific problem, typically meaning one that is at least as smart as a typical human.

While narrow AI may outperform humans at whatever its specific task is, for instance playing chess or solving mathematical equations, AGI would outperform humans at nearly every cognitive task. The ultimate hypothetical goal is achieving superintelligence (ASI) which is far surpassing that of the brightest and most gifted human minds. Due to recursive self-improvement, superintelligence is expected to be a rapid outcome of creating artificial general intelligence. The diagram below by Ziyad Mohammed 2018/2019 outlines the progression plan of AI.

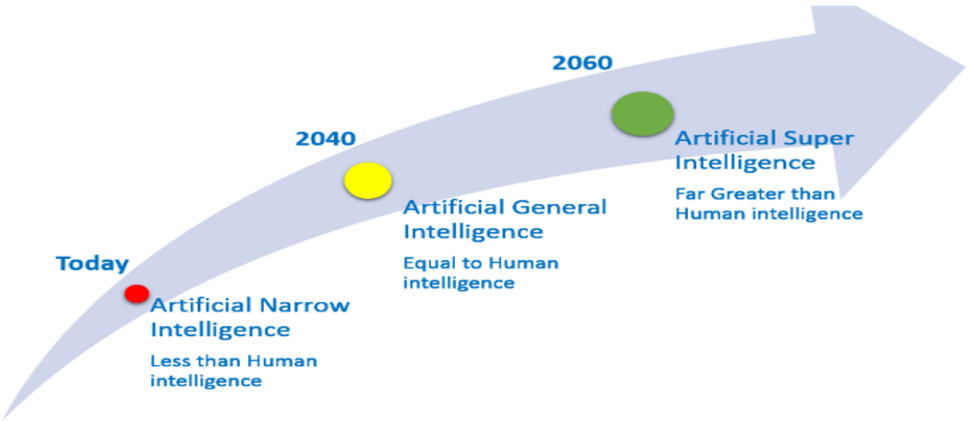



Fig. 1

In addition, Hanif Khan, September 2021 describes the functions of ANI, AGI and ASI are stated as follows: ANI specializes in one area and solves one problem (Machine Learning), AGI refers to a computer that is as smart as a human being (Machine Intelligence) and ASI refers to a machine that is far smarter than a human in a given field. Its intellect higher than that of a human (Machine Consciousness). Hanif Khan, September 2021 provides more diagrams that give a visual view of what each AI types does.

### 1.2.2 Type 2-based on functionalities

**Purely Reactive**

Reactive machines are basic in that they do not store 'memories' or use past experiences to determine future actions. They simply perceive the world and react to it. IBM's Deep Blue, which defeated chess grandmaster Kasporov, is

a reactive machine that sees the pieces on a chess board and reacts to them. It cannot refer to any of its prior experiences, and cannot improve with practice.

**Limited Memory**

Limited Memory machines can retain data for a short period of time. While they can use this data for a specific period of time, they cannot add it to a library of their experiences. Many self-driving cars use Limited Memory technology: they store data such as the recent speed of nearby cars, the distance of such cars, the speed limit, and other information that can help them navigate roads.

**Theory of Mind**

Psychology tells us that people have thoughts, emotions, memories, and mental models that drive their behaviour. Theory of Mind researchers hope to build computers that imitate our mental models, by forming representations about the world, and about other agents and entities in it. One goal of these researchers is to build computers that relate to humans and perceive human intelligence and how people's emotions are impacted by events and the environment.

**Self-Awareness**

Self-aware machines are the stuff of science fiction, though many AI enthusiasts believe them to be the ultimate goal of AI development. Even if a machine can operate as a person does, for example by preserving itself, predicting its own needs and demands, and relating to others as an equal, the question of whether a machine can become truly self-aware, or 'conscious', is best left for philosophers.

### 1.3 Machine learning

Machine learning is the ability of machines to automatically learn from data and algorithms, and is one of the more demanding branches of artificial intelligence. Paul Blocchi, 2023 defines Machine Learning as a subset of Artificial Intelligence that involves the development of algorithms and statistical models that enable computers to learn from data and improve their performance over time without being explicitly programmed. Machine learning improves performance using past experiences and can make decisions without being specifically programmed to do so. The process starts with historical data collection, like instructions and direct experience, so that logical models can be built for future inference. Output accuracy depends on data size – a larger amount of data will build a better model, which in turn increases its accuracy. Machine learning algorithms are classified into three types:

- **Supervised learning.** Machines are trained with labeled data to predict the outcome.
- **Unsupervised learning.** Machines are trained with unlabeled data, with the model extracting information from the input to identify features and patterns, so it can generate an outcome.
- **Reinforcement learning.** Machines learn through trial and error, using feedback to form actions.

### 1.3.1 How does Machine Learning work

A machine learning system builds prediction models, learns from previous data, and predicts the output of new data whenever it receives it. The amount of data helps to build a better model that accurately predicts the output, which in turn affects the accuracy of the predicted output. When we have a complex problem in which we need to make predictions. Instead of writing code, we just need to feed the data to generic algorithms, which build the logic based on the data and predict the output. Our perspective on the issue has changed as a result of machine learning. The Machine Learning algorithm's operation is depicted in the following block diagram (source:www.javatpoint.com/machine-learning, 2024).:

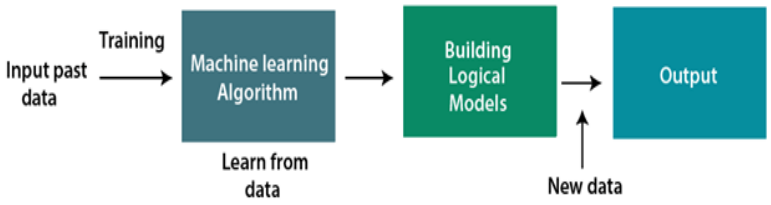



Fig.2

### 1.3.2 What is Supervised Learning?

Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross-validation process. Supervised learning helps organizations solve for a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox. This according to the source "What is Supervised Learning". (2024). Supervised Learning. Available at: https://www.ibm.com (Accessed:15th January 2024). In supervised learning, sample labeled data are provided to the machine learning system for training, and the system then predicts the output based on the training data. The system uses labeled data to build a model that understands the datasets and learns about each one. After the training and processing are done, we test the model with sample data to see if it can accurately predict the output. The mapping of the input data to the output data is the objective of supervised learning. The managed learning depends on oversight, and it is equivalent to when an understudy learns things in the management of the educator. Spam filtering is an example of supervised learning as described by the source "Supervised Machine Learning". (2024) Supervised Machine Learning. Available at: https://www.javatpoint.com (Accessed: 15 January 2024)

### 1.3.3 How supervised learning works?

In supervised learning, models are trained using labelled dataset, where the model learns about each type of data. Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output according to the source "Machine Learning". (2024) Machine Learning. Available at https://www.javatpoint.com (Accessed:16th January 2024).

The image below is a visual representation of supervised machine learning

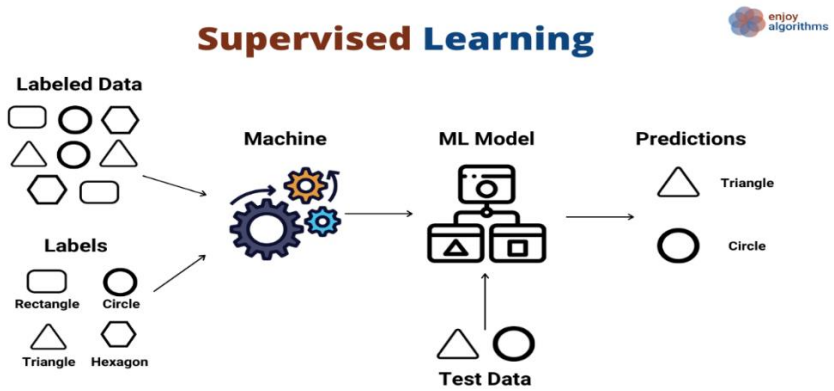



Fig.3

### 1.3.4 Categories of Supervised Machine Learning

There are two categories of supervised machine learning: Regresion and Classification. In our study regression is our main area of interest.

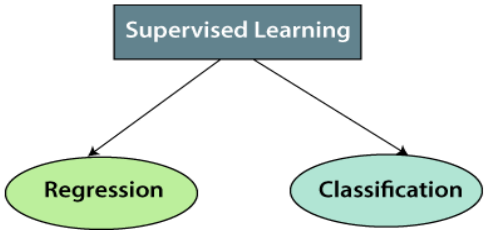



Fig.4

**Regression**

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends. Examples are Linear Regression, Regression Trees, Non-Linear Regression, Bayesian Linear Regression, Polynomial Regression. For the purpose of this study we shall concentrate on Linear Regression as a statistical method we shall use to implement our solution in supervised machine learning.

**Features of Linear Regression**

The following are the features of linear regression as accessed. Regression Analysis. (2024) Regression Analysis. Available at: www.javatpoint.com (Accessed: 15 January 2024).

- Linear regression is a statistical regression method which is used for predictive analysis.
- It is one of the very simple and easy algorithms which works on regression and shows the relationship between the continuous variables.
- It is used for solving the regression problem in machine learning.
- Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), hence called linear regression.
- If there is only one input variable (x), then such linear regression is called **simple linear regression**. And if there is more than one input variable, then such linear regression is called **multiple linear regression**.
- The relationship between variables in the linear regression model can be explained using the below image. Here we are predicting the possibility of defaulting/not paying ground rent bills by property owners on the basis of **the unpaid years which are increasing**.

**1.3.5 Implementing the Supervised Machine Learning Solution**

There are several platforms that offer machine learning services such as Anaconda Navigator. In this study, I will concentrate on Anaconda Navigator Jupyter Notebook to build, train and test a machine learning algorithm that uses linear regression statistical method. According to the AWS Developer Guide, 2024, data scientists and developers can quickly build, train, and deploy ML models into a production-ready hosted environment. The input dataset can be in CSV format. The following diagram in fig.5 illustrates the process of building, training, testing and deploying a machine model.

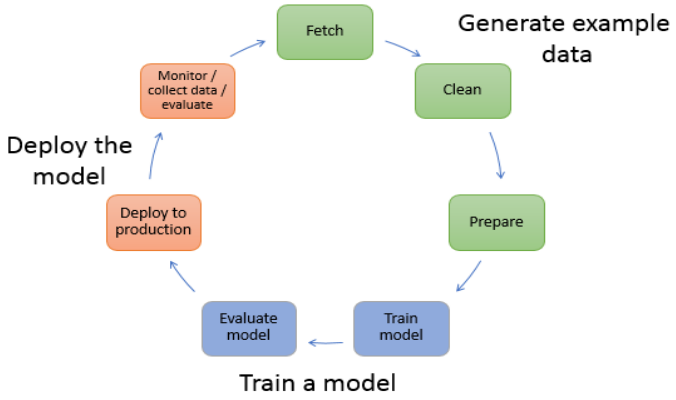



Fig.5

## 2. Method

This phase of our study will demonstrate the process of building, training and testing a supervised machine learning model. The following are the process involved: data importation; data splitting; building the machine model, train and test; and evaluating the model.

✓ **Data Importation**

- This stage involves importing the libraries needed to import the data and begin the process of building the machine learning model.

```
import pandas as pd
import numpy as py
ground_rent_data=pd.read_csv(r"G:\Linear Regression\Lusaka Province Ground Rent Fees.csv")
ground_rent_data
```

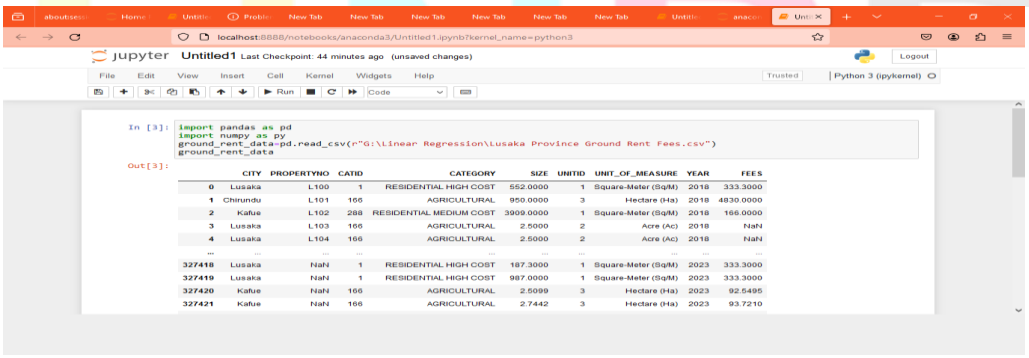


Fig.6

- The next stage requires us to remove rows from the dataframe with NaN or null values.

```
ground_rent_data.dropna(subset=['PROPERTYNO'], inplace=True)
ground_rent_data.dropna(subset=['FEES'], inplace=True)
ground_rent_data
```

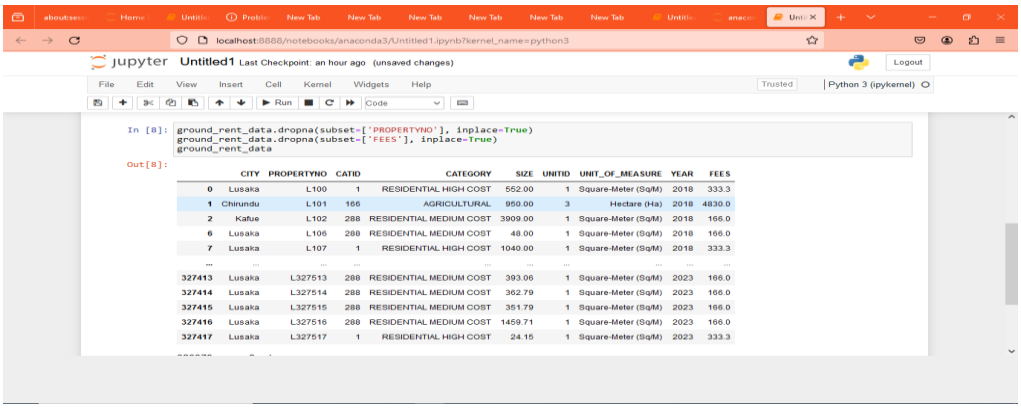



Fig.7

- The next stage required is to identify the fields that we are going to use in carrying out our work.

```
ground_rent_data= ground_rent_data [["CATID","UNITID","SIZE","FEES"]]
ground_rent_data
```

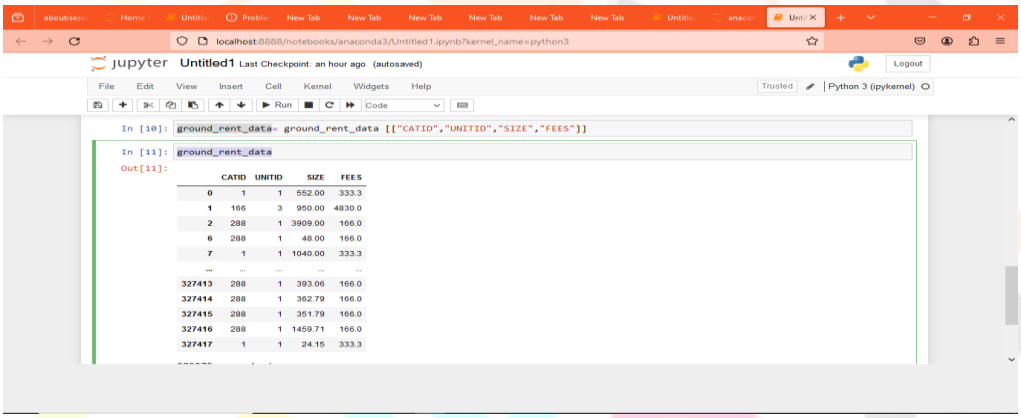



Fig.8

- The next stage is to identify the x variable (independent) and y variable (dependent)

```
x= ground_rent_data.drop(["FEES"], axis=1).values
y= ground_rent_data ["FEES"].values
```

✓ **Data Spliting: Train and Test Data**

- This part involves splitting the dataframe into Training and Test datasets The training data is used for training the model during its build stage, while the test data is used for testing the model after it has been built. We start by importing the necessary library for splitting the dataframe into training and testing.

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

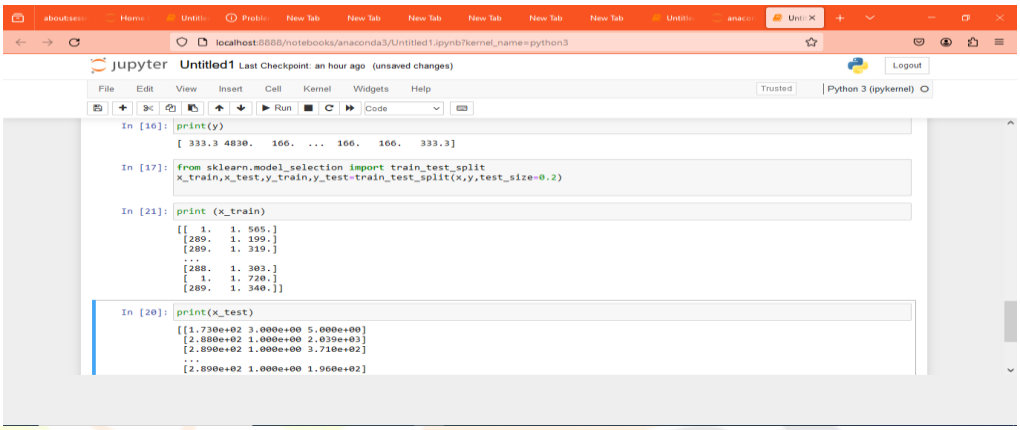



Fig.9

✓ **Building the Machine Model**

- This phase we shall build the machine model, train and test it on the dataframe split. Further we are going to run the predict function to predict future values or fees (see fig.12).

```
from sklearn.linear_model import LinearRegression

reg=LinearRegression()
reg.fit(x_train,y_train)
y_pred=reg.predict(x_test)
print(y_pred)
```

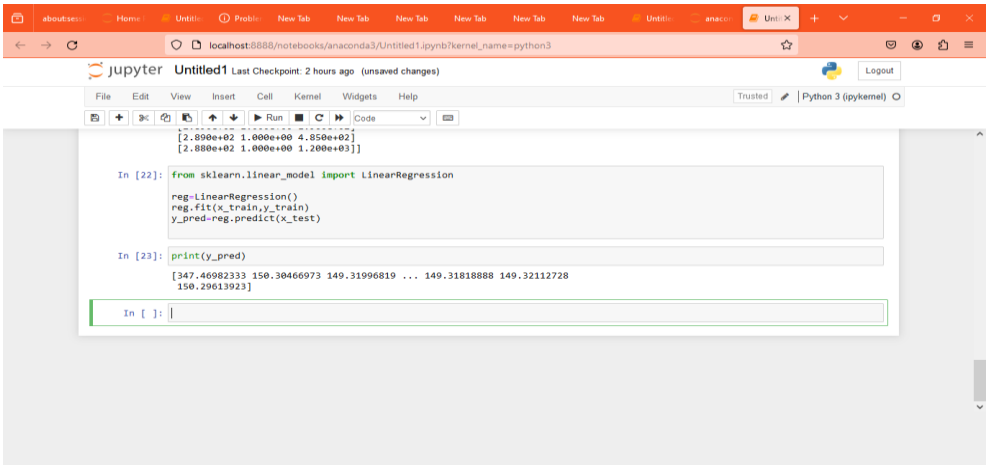


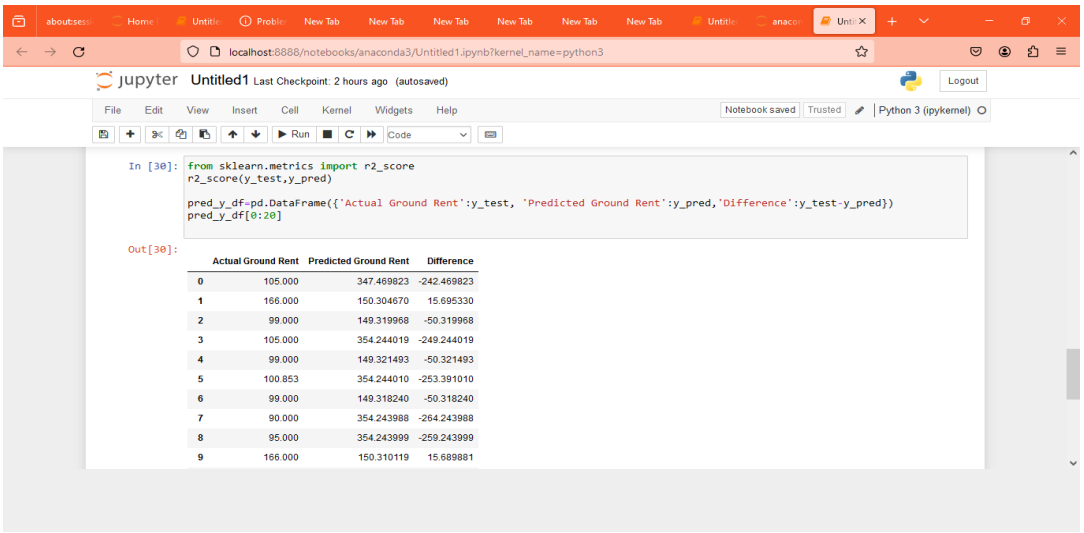

Fig.10

**Note:** The training stage can be repeated multiple times until the performance of the machine learning model improves.

✓ **Evaluating the Machine Model**

- The evaluating part requires that the predicted values are compared against the actual values.

```
from sklearn.metrics import r2_score
r2_score(y_test,y_pred)
pred_y_df=pd.DataFrame({'Actual Ground Rent':y_test, 'Predicted Ground Rent':y_pred,'Difference':y_test-y_pred})
pred_y_df[0:20]
```




| | Actual Ground Rent | Predicted Ground Rent | Difference |
|---|---|---|---|
| 0 | 105.000 | 347.469823 | -242.469823 |
| 1 | 166.000 | 150.304670 | 15.695330 |
| 2 | 99.000 | 149.319968 | -50.319968 |
| 3 | 105.000 | 354.244019 | -249.244019 |
| 4 | 99.000 | 149.321493 | -50.321493 |
| 5 | 100.853 | 354.244010 | -253.391010 |
| 6 | 99.000 | 149.318240 | -50.318240 |
| 7 | 90.000 | 354.243988 | -264.243988 |
| 8 | 95.000 | 354.243999 | -259.243999 |
| 9 | 166.000 | 150.310119 | 15.689881 |

Fig.11

The screen shot in fig.11 indicates the differences that exist between the actual ground rent fees and the predicted ground rent fees.

## 3. Results and Discussion

The tests carried out on the machine model produced the predicted results in the table (fig.12) below , which are compared against the actual results. Noted that the machine model can be tested repeatedly until its performance improves.

| SNO | Actual Ground Rent | Predicted Grount Rent | Difference |
|---|---|---|---|
| **1** | 166.000 | 150.304670 | 15.695330 |
| **2** | 99.000 | 149.319968 | -50.319968 |
| **3** | 105.000 | 354.244019 | -249.244019 |
| **4** | 99.000 | 149.321493 | -50.321493 |
| **5** | 100.853 | 354.244010 | -253.391010 |
| **6** | 99.000 | 149.318240 | -50.318240 |
| **7** | 90.000 | 354.243988 | -264.243988 |
| **8** | 95.000 | 354.243999 | -259.243999 |
| **9** | 166.000 | 150.310119 | 15.689881 |
| **10** | 333.300 | 428.051659 | -94.751659 |
| **11** | 166.000 | 150.297298 | 15.702702 |
| **12** | 105.000 | 354.244019 | -249.244019 |
| **13** | 99.000 | 149.324279 | -50.324279 |
| **14** | 166.000 | 150.331197 | 15.668803 |
| **15** | 580.000 | 513.922451 | 66.077549 |

Fig.12

This study was carried using multi-linear regression algorithm used in statistical prediction. The following are the findings of the research.

- ✓ **Multi-Linear Regression Statistical Algorithm**

Multi-Linear Regression Algorithm is an efficient, effective and appropriate tool for demonstrating the relationship between independent and dependant variables. Multi-Linear Regression Algorithm enables the researcher to work with multiple independent variables, sometimes called predictors. The reason is that in a real-world experience ground rent fees are determined by multiple factors in this case land size, category and unit of measurement in this case square metres, acres and hectares.

- ✓ **Machine Learning can process very Large datasets**

Supervised Machine Learning models have the ability to process very large datasets efficiently and effectively. This means that records in their millions can be processed by the model in the shortest possible time. These records can be from various data sources such as traditional databases, data warehouses, text files, csv files and images, audio and videos. This eliminated the need for humans to analyse data amounting to millions which may take a very long time.

- ✓ **Predicted Values Vs Actual Values**

Predicted values, are values predicted by the machine learning model. The predicted values are expected to be accurate or closer to the actual values extracted from the data source. The difference between the predicted and actual values is called the '**error**' and it is a normal occurance. The '**error**' can be reduced by repeatedly training the model until its performance improves.

- ✓ **Clean and Accurate Data**

It important to note that clean and accurate data, are essential to have an effective and efficient machine learning model. In the case of our study, records with columns such as property number and fee's with null values had to be identified and removed. It is important to remove such records as processing may stop mid-way and error's thrown. Most machine learning models do not work with null values, therefore during this research the data had to be cleaned or identify records with columns that had null values and removed.

- ✓ **Independent and Dependent Variables**

It is important to identify the correct independent and dependent variables for the machine learning model to make accurate predictions. As stated earlier the independent variables in our research are land size, category and unit of measurement, while the dependent variable is the fees column. Independent variables determine the dependent variable. The independent variables are predictors, while the dependent variable is a target. In the research when land size is high and in hectares as unit of measurement then the fees will tend to be high as well.

## 4. Conclusion

The future management of ground rent fees can be done through using artificial intelligence powered machine learning models. These machine learning models have the ability to improve performance of an organization effectively and efficiently. The tools as can be seen are readily available on the internet and the popular tools being Anaconda Navigator Jupyter Notebook, Amazon Web Services tools and Amazon cloud service. The ministry could leverage on these tools by simply subscribing to Amazon or other plaforms and use them to develop as many machine learning models to help solve many problems that may be routine in nature in order to minimize on the workload placed on employees.

## Acknowledgement

I wish to acknowledge and appreciate my lecturers at DMI St Eugene University for the time and effort rendered in introducing me to Artificial Intelligence and Machine Learning areas of technology where I wish to pursue further studies and develop future solutions. Further am grateful to my family for the support received during the last 2 years of my Master of Science Degree study.

## References

1. Ministry of Justice, Government of the Republic of Zambia (1975), Land Act of 1975.
2. Regression Analysis. (2024) Regression Analysis. Available at: www.javatpoint.com (Accessed: 15 January 2024).
3. Ziyad Mohammed Saleh. (2019) Artificial Intelligence Definition, Ethics and Standards. The British University of Egypt. WATHI.
4. Paul Blocchi. (2023) 6 Branches Artificial Intelligence. Available at: https://www.linkedin.com/in/paulblocchi/ (Accessed:10 th January 2024).
5. Hanif Kan. (2021) Different Types of AI Systems. Ha Noi. University of Science and Technology.
6. Supervised Machine Learning. (2024) Supervised Machine Learning. Available at: https://www.javatpoint.com (Accessed: 15 January 2024)
7. Machine Learning. (2024) Machine Learning. Available at https://www.javatpoint.com (Accessed:16th January 2024).
8. Amazon Web Services (2023), AWS Step Functions Developer Guide.