**JENGA SCHOOL OF DATA SCIENCE AND AI**

**ASSESSING OFFENDERS’ RISK: PREDCITING RECIDIVISM USING MACHINE LEARNING**

**FELISTA MOGIRE**

**SUPERVISOR: DR. LAWRENCE NDERU**

**AUGUST 2022**

**1.0 OVERVIEW**

Recidivism is the return to criminal acts by an offender after release from a correctional system. Judicial systems around the world use predictions of an offender’s risk of recidivating to partly inform decisions such as sentencing, release, bail and bond terms. The national average of reoffending in Kenya is 47 %. The Probation and After Care Service department is tasked with providing advisory reports to court and other agencies for the purpose of determining bail and bond terms and for decision-making on sentencing and release of offenders. One of the factors the program takes into consideration in their advisories is an offender’s risk of recidivating.

In Kenya, like in many other countries, clinical judgements, criminal history, record reviews, risk assessment tests, statistical methods and other forensic techniques are used to predict recidivism. The predicted risk level helps to determine offender’s release, length of sentences, bail amounts and also rehabilitation strategies. Accuracy of the predicted risk is very important as it ensures that courts optimally produce sentences that are fair but also protect the society. While existing methods of predicting recidivism work, they have limitations which can affect accuracy. In addition they can also be time consuming.

Machine learning has the potential to improve the predictive ability for recidivism over classical statistical methods. Classical statistical methods of predicting recidivism largely assume a linear relationship between the risk of reoffending and static variables (e.g. place of birth, gender, age) or dynamic variables (e.g. marital status, level of Education) of the offender. Since recidivism is a complex problem it is possible that the relationship between some variables and recidivism is non-linear and therefore this factor lowers the accuracy of classical statistical methods. Machine learning methods are non-parametric and can be used to evaluate non- linear relationships and perform feature selection. This project proposes to apply supervised learning classification method of machine learning to determine a set of optimal predictors for recidivism.

With supervised learning the project will create a model that will predict the outcome (will recidivate or not) based on the relationships among the predictor (input) variables. An ideal dataset for this project will contain a variety of information including the offenders’ ages, gender, place of residence, level of education, type of crime committed, whether first time or repeat offender, employed or not, ethnicity, violent or non-violent crime and much more. The approach will be to

* 1. **CRISP-DM METHODOLOGY**

The project will follow the CRISP-DM methodology of implementation. CRISP-DM stands for cross-industry processes for data mining. The model was published in 1999 to standardize data mining processes across industries. CRISP-DM provides a template to ensure you have considered all of the different aspects specific to your data science project. This process model has six phases that naturally describe the data science life cycle. The six phases are: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment.

**2.0 BUSINESS UNDERSTANDING**

2.1 Business Objective

This project aims to apply machine learning algorithms in recidivism risk assessment. The assessments will be used in trials and sentencing decisions. Assessing recidivism risk using machine learning will produce more accurate analyses that are also less biased than those of human beings. Moreover, decreasing budgets, lack of resources and personnel has called for more systematic and objective methods of processing offenders. In the United States, predictive algorithms have been used extensively in the criminal justice system to predict where crimes will most likely occur, who is most likely to commit a crime, who is likely to fail to appear at their court hearing, and who is likely to reoffend at some point in the future. Kenya on the other hand has yet to apply predictive algorithms in the criminal justice system.

This project seeks to demonstrate to the Probation & Aftercare Service and the Judiciary of Kenya that the efficacy and accuracy of recidivism risk assessment can be improved greatly. By having a tool that can significantly reduces the time used in assessment and with a high accuracy, the Kenyan government can save on resources. Also, the Judiciary of Kenya has always grappled with the challenge of slow processing of criminal cases and increase in backlog. Automated recidivism assessment can go a long way in assisting the judicial system make decisions quicker and hence improving its efficacy. Automated recidivism analysis can also bring about fairness in the system by removing bias which comes about when human beings asses.

2.2 Data Mining Goals

i. Identifying a set of optimal predicators (variables), whether static or dynamic for recidivism from past criminal data.

ii. Building a model using the past criminal data to predict the likelihood of an offender recidivating.

iii. Classifying an offender either as one who will recidivate or not

iv. For the offenders who are classified as ‘will recidivate’, assigning the offender a rank for how high or low the risk is.

**3.0 DATA UNDERSTANDING**

The ideal dataset for this project will be one that is obtained from the Kenyan Judicial system for at least one of the counties. The project is in pursuit of this dataset but as there might be a delay in acquiring it, another dataset for the IOWA state of the United States has been obtained from Kaggle for piloting.

3.1 The Dataset

This dataset reports whether an offender is re-admitted to prison or not within three years from being released from prison in Iowa. It contains entries of a 3-year time period after an initial release from prison, among offenders serving a prison term in the state of Iowa, US between 2010 and 2015, with recidivism follow-up between 2013 and 2018. The dataset contains 26021 records.

The dataset has the following columns:

|  |  |  |
| --- | --- | --- |
| Column Name | **Data Type** | **Description** |
| Fiscal Year Released | Number | Fiscal year (year ending June 30) for which the offender was released from prison. |
| Recidivism Reporting Year | Number | Fiscal year (year ending June 30) that marks the end of the 3-year tracking period. For example, offenders exited prison in FY 2012 are found in recidivism reporting year FY 2015. |
| Main Supervising District | Text | The Judicial District supervising the offender for the longest time during the tracking period. |
| Release Type | Text | Reasoning for Offender's release from prison. |
| Race-Ethnicity | Text | Offender's Race and Ethnicity |
| Age At Release | Text | Offender's age group at release from prison. |
| Sex | Text | Gender of the offender |
| Offense Classification | Text | Maximum penalties: A Felony = Life; B Felony = 25 or 50 years; C Felony = 10 years; D Felony = 5 years; Aggravated Misdemeanor = 2 years; Serious Misdemeanor = 1 year; Simple Misdemeanor = 30 days |
| Offense Type | Text | General category for the most serious offense for which the offender was placed in prison. |
| Offense Subtype | Text | Further classification of the most serious offense for which the offender was placed in prison. |
| Return to Prison | Text | No = Did not return to prison within the three year tracking period; Yes = Admitted to prison for any reason within the three year tracking period |
| Days to Return | Number | Number of days it took before the offender returned to prison. |
| Recidivism Type | Text | Indicates the reason for return to prison. |
| New Offense Classification | Text | New conviction maximum penalties: A Felony = Life; B Felony = 25 or 50 years; C Felony = 10 years; D Felony = 5 years; Aggravated Misdemeanor = 2 years; Serious Misdemeanor = 1 year; Simple Misdemeanor = 30 days |
| New Offense Type | Text | General category for the new conviction while the offender is out of prison. |
| New Offense Sub Type | Text | Further classification of the new conviction. |
| Target Population | Text | The Department of Corrections has undertaken specific strategies to reduce recidivism rates for prisoners who are on parole. |