

---

# **MACHINE LEARNING: APPROACHES TO TAX COMPLIANCE RISK MANAGEMENT**

Amos Kiptui

Data Governance - Kenya Revenue Authority

# GLOBALIZATION JOURNEY



- Exploration
- World: Large to Medium
- Driving Force: Muscle (Power, Global Integration)

Globalization  
1.0

Globalization  
2.0

- World: Medium to Small
- Driving Force: Multinational Companies (Markets & Labor, Communication, Global Economy)

- World: Small to Tiny
- Driving Force: Collaboration & Automation

Globalization  
3.0

Globalization  
4.0

- World: Flat
- Driving Force: Knowledge Economy

# TAX ENVIRONMENT: AFRICAN PERSPECTIVE



Growing Informal Sector



Limited Resources to Sustain Compliance Checks (Taxpayer-to-Taxpayer)



Tax Administrations are grappling with changing taxpayer behaviour and new technologies that increase compliance risks.



Low Tax Morale (Tax Culture, Low Voluntary Compliance Levels)

# TAX COMPLIANCE RISK MANAGEMENT

## MACHINE LEARNING

Compliance risk management refers to the process of identifying, assessing, ranking and treating compliance risks.

Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.



# COMPLIANCE RISKS



Registration



Declaration



Payment



Reporting





# WHY RISK MANAGEMENT?

## FOR DATA



"War is ninety percent information."

Napoleon Bonaparte



"I never guess. It is a mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

Arthur Conan Doyle

## FOR INTUITION



"You can't connect the dots looking forward; you can only connect them looking backwards. So you have to trust that the dots will somehow connect in your future. You have to trust in something- your destiny, gut, life, karma, whatever. This approach has never let me down, and it has made the difference in my life."

Steve Jobs

# ...WHY RISK MANAGEMENT?

- **Use of Data vs Intuition** - tax administrations especially in developing countries are identifying risks based on tax officers' experience and not based on data and this often leads to wrong conclusions.
- **Compliance Risk Management Tools** - the current tools of assessing risks are not accurate and prone to human error.
- **Ever-Changing Taxpayer Behavior** - the need to develop and continuously update risk registers based on emerging risks is time consuming and resource intensive.
- **Centralization of Risk Management** - there is a need to centralize risk management based on standard criteria and models from a corporate level.



# DEVELOPING THE MODEL

- Step 1: Choose Target Group/Population
- Step 2: Define Target Parameters
- Step 3: Define Risk Indicators
- Step 4: Define the Weight and Scores of Risk Indicators
- Step 5: Developing the Regression Model



# DEVELOPING THE MODEL: EXAMPLE

## ▪ **Step 1: Choose Target Group/Population**

*A target population is a hard delimiter (does not change) that describes the specific group for which the analysis is focused on.*

**Alcoholic beverages**

# DEVELOPING THE MODEL: EXAMPLE

## ▪ **Step 2: Choose Target Parameter**

*A target parameter is a soft delimiter (can vary from time to time) that describes the overall objective of carrying out the selection analysis.*

**Tax Fraud: Under declaration of quantity of goods.**

# DEVELOPING THE MODEL: EXAMPLE

## ▪ **Step 3: Define Risk Indicators**

*Risk indicators are criteria that serve as a tool to target and gauge potential risks within the target population.*

- ✓ **Variance in the gross weight of goods declared versus the sector average quantity**
- ✓ **Variation in container gross weight and net weight**

# DEVELOPING THE MODEL: EXAMPLE

## ▪ Step 4: Define the Weight and Scores of Risk Indicators

*Risk indicators are criteria that serve as a tool to target and gauge potential risks within the target population.*

### ✓ **Variance in the gross weight of goods declared versus the sector average quantity**

- The gross weight should not be below 10% of the sector average of 17,000 kgs
- 10 to 20%=3, >20-30%= 4, > 30% = 5

### ✓ **Variation in container gross weight and net weight**

- The net weight should not exceed 10% of the gross weight
- 10-30%= 3, > 30% to 50% = 4, >50%=5

# DEVELOPING THE MODEL: EXAMPLE

## ▪ Step 4: Developing the Model

### ✓ **Weights**

| Risk             | Rank | Weight |
|------------------|------|--------|
| Registrati<br>on | 4    | 10     |
| Declaratio<br>n  | 3    | 30     |
| Payment          | 2    | 20     |
| ✓ Reporting      | 1    | 40     |

Score = risk points \* weight

$$y = \text{score1} + \text{score2}$$

*Predict chances of a taxpayer being non-compliant (risk levels) based on characteristics of target group.*



# APPLICATION OF MACHINE LEARNING IN TAX

- Issuance of Tax Compliance Certificate
- Processing of Refund Claims
- Monitoring Taxpayer Behavior – chances of non-compliance (apply treatment as early as possible)
- Fraud Detection – Post Clearance Audit (Customs)

# CHALLENGES

- Poor Quality Data
- Dynamic Tax Environment – changing taxpayer behavior and new risks

# RECOMMENDATIONS

- Use 3<sup>rd</sup> party data to enhance model
- Start small – key target population and indicators first
- Enhance data quality



# Thank You



amoskiptui



me@amoskiptui.co.ke

[www.amoskiptui.co.ke](http://www.amoskiptui.co.ke)