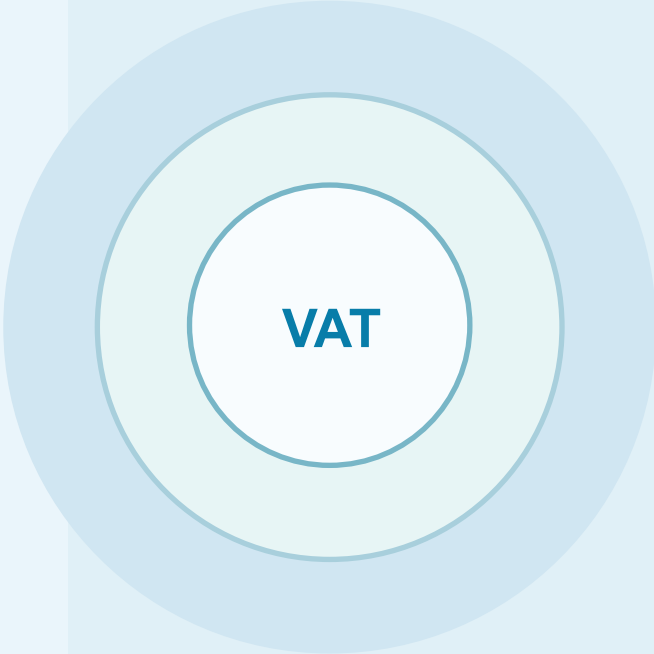
GEORGIA REVENUE SERVICE

IOTA Workshop · SOFIA 2026

Enhancing Automated VAT Refund System with Machine Learning



Georgia Revenue Service

SOFIA · 2026

Objectives of the Reform

Two priorities guiding the VAT refund reform

01

Simpler, timely access

Enable taxpayers to access their own funds through a simpler and more timely refund process.

02

Reduce accumulated overpayments

Limit the build-up of overpaid VAT amounts and mitigate the risks associated with them.

Automated VAT Refund System

INPUT

VAT

Taxpayers

VAT Returns

100% submitted

RISK ASSESSMENT

Module 1 | Module 2

Module 3 | Module 4

Module 5 | Module 6

Rule-based screening

ROUTING

Risk assessment

>90% **VAT (green) refund card**

Payment order → Bank

<10% **Departmental review**

Monitoring / Audit department

Rule-Based Risk Assessment

Illustrative risk indicators and the corresponding scoring logic

Compliance History

RISK INDICATOR	RULE	PTS
Late payment	> 3	1
Late filing	> 2	2
Late registration	Yes/No	0
Fines	> 2	1
Frequent declaration adjustments	> 8	1

Company Management

RISK INDICATOR	RULE	PTS
Director under age 20	Age < 20	2
Elderly director	…	1
Non-resident founder	…	2
Founder created another indebted company	…	3
Unreachable / uncooperative contact	…	1

Operations

RISK INDICATOR	RULE	PTS
Frequent credit declarations	…	1
Invoices involving fraudulent taxpayers	…	3
High output VAT without matching imports / purchases	…	3
Customs violations	…	0
Additional indicators	…	…

SCORING EXAMPLE

Taxpayer score	Maximum possible	Relative score	Declared amount	Final score
21	75	0.28	10,000	2,800

Limitations of the Rule-Based Approach

1	**High maintenance**	Hundreds of rules make systems hard to maintain and slow to adapt to changing taxpayer or trading-partner behavior.
2	**False positives**	Low-risk activity may be flagged as high risk and it can be difficult to identify which rules caused the error.
3	**Frequent manual updates**	Changes in tax legislation or declaration forms require identifying and revising all related rules.
4	**Difficult to improve accuracy**	Poor performance may reflect missing rules as well as existing rules that produce misleading results.
5	**Limited use of available data**	Including thousands of data items becomes increasingly complex as the number of variables grows.

Refining the Automated VAT Refund System

INPUT

VAT

Taxpayers

VAT Return

100% submitted

RISK ASSESSMENT

Module 1 | Module 2 | Module 3

Module 4 | Module 5 | Module 6

RANDOM | ML

Additional risk-assessment inputs

ROUTING

Risk assessment

>90% **VAT (green) refund card**

Payment order → Bank

<10% **Departmental review**

Monitoring / Audit department

ML Implementation Process

1

Define population

Define the taxpayer population for data collection, feature calculation and analysis.

2

Feature engineering

Currently there are 73 taxpayer features, each with its own logic.

3

Account periods

Use four consecutive periods to evaluate taxpayers and features over time.

4

Random Forest training

Train the model in Python using a Pipeline with a defined process chronology.

5

Scoring & data store

Score new records, return results to the data store, and prepare top-risk taxpayers for case opening.

Four consecutive account periods

20.10.2025 - 20.10.2026

20.10.2024 - 20.10.2025

20.10.2023 - 20.10.2024

20.10.2022 - 20.10.2023

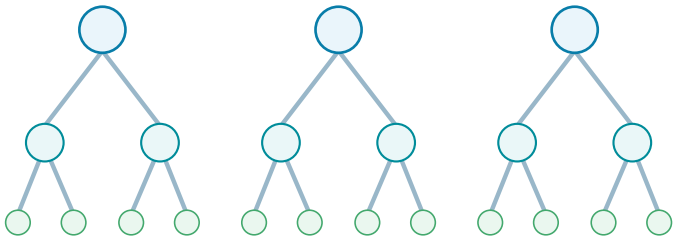
Model Training & Scoring

1 Feature Store	Extract the defined features and prepare the dataset for the modelling pipeline.
2 Upload & training	Upload the data and train Random Forest on the learning subset (e.g. 700 of 1,000 records).
3 Refit	Retrain the selected model on the full dataset (e.g. all 1,000 records).
4 New model version	Create the new model version and assign scores to new taxpayer records.
5 Results & case opening	Return results to the data store. Send top-risk taxpayers to Oracle SQL procedures for case opening (This process is still under development and not yet functional).

Random Forest

Why this model is useful for VAT risk assessment

An ensemble of decision trees



COMBINED PREDICTION

01 Stability

Ensemble learning helps reduce overfitting.

02 Ease of use

Relatively straightforward to work with.

03 Flexibility

Supports both classification and regression.

04 Feature importance

Helps identify influential predictors.

Challenges of Machine Learning

01	**Institutional readiness**	Tax officials and the public may be reluctant to trust automated risk assessment or auditing systems.
02	**Accountability**	Assigning responsibility for automated decisions can be legally complex.
03	**Data quality and completeness**	Reliable, complete data are a central requirement.
04	**Skills and infrastructure**	Capability is needed to manage large datasets with diverse structures and formats.
05	**Security and confidentiality**	Protecting information and managing confidentiality concerns remain essential.

GEORGIA REVENUE SERVICE

Thank you

Georgia Revenue Service

IOTA Workshop · Sofia 2026