Ecole Nationale Supérieure d'Ingénieurs de Tunis

CIMF
CENTRE INFORMATIQUE DU MINISTÈRE DES FINANCES

Design, and develop a self-learning and smart machine to analysis the tax and accounting data masses to detect tax fraud and build predictive models in order to help detect in advance fraudsters.

Realised by: Sameh HABBOUBI

Encadré par :
Mrs Wahiba BEN FREDJ (Entreprise Supervisor)
Mr Wajdi GARALI (ENSIT Supervisor)

Introduction

Ministry of Finance data wants to automate and target high income verification missions in terms of money and time to select taxpayers who have the highest recovery amount.

Problematic

- Large data amounts.
- Decrease human intervention .
- Streamline tasks.
- Automating control spot.

Solution

Machine Learning algorithms
+
Data preprocessing
=
Intelligent Machine for Financial Analysis

Methodology

CRISP-DM methodology: provides a structured approach to planning a data mining project.
CRISP-DM stands for cross-industry process for data mining.

Results

Machine learning algorithms prediction model best results:

	precision	recall	f1-score	support
[0,1000dt[	0.45	0.34	0.39	134
[1000dt,3500dt[	0.43	0.47	0.45	178
[3500dt,10000dt[	0.38	0.44	0.41	135
[100000dt, 300000dt[	0.46	0.54	0.50	100
[300000dt, 60000dt[	0.40	0.15	0.22	26
[600000dt, Inf[	0.71	0.47	0.57	32
avg / total	0.44	0.43	0.43	605

	precision	recall	f1-score	support
[0,1000dt[	0.43	0.17	0.24	54
[1000dt,3500dt[	0.36	0.29	0.32	73
[3500dt,10000dt[	0.41	0.39	0.40	98
[10000dt, 30000dt[	0.32	0.47	0.38	113
[30000dt, 60000dt[	0.25	0.10	0.14	59
[60000dt, Inf[	0.44	0.63	0.52	82
avg / total	0.37	0.37	0.35	479

Conclusion

Using these machine learning models, we can predict fraudsters and their tax adjustment amount.