

Dominik Zappe

Diplomová práce

Informatika a její specializace
Zpracování přirozeného jazyka
2024/2025Vedoucí práce:
doc. Ing. Jan Pospíšil, Ph.D.
Katedra matematiky FAV ZČUDetekce anomálií v datech
z knih limitních objednávek

Abstrakt

Moderní finanční trhy jsou rychlé, komplexní a stále častěji se stávají cílem sofistikovaných forem manipulace. Tato práce se zaměřuje na detekci anomálií v časových řadách odvozených z knih limitních objednávek s cílem rozpoznat manipulativní chování zvané spoofing. Vzhledem k absenci anotovaných dat jsou použity metody strojového učení bez učitele aplikované na reálná historická data. V práci je implementováno šest metod – izolační les, lokální faktor odlehlosti, jednotřídní SVM, plně propojený autoenkodér, konvoluční autoenkodér a transformer autoenkodér. Modely jsou evaluovány pomocí méně známých metrik Excess Mass a Mass Volume, přičemž nejlépe si vedou modely izolační les a transformer. Kombinací nejvýkonnějších modelů vznikl robustní nástroj schopný odhalit podezřelé chování bez ruční anotace. Navržené řešení efektivně identifikuje rizikové oblasti pro následnou expertní analýzu a představuje tak praktický přínos pro detekci nelegálních praktik na finančních trzích.

Hlavní aspekty realizace

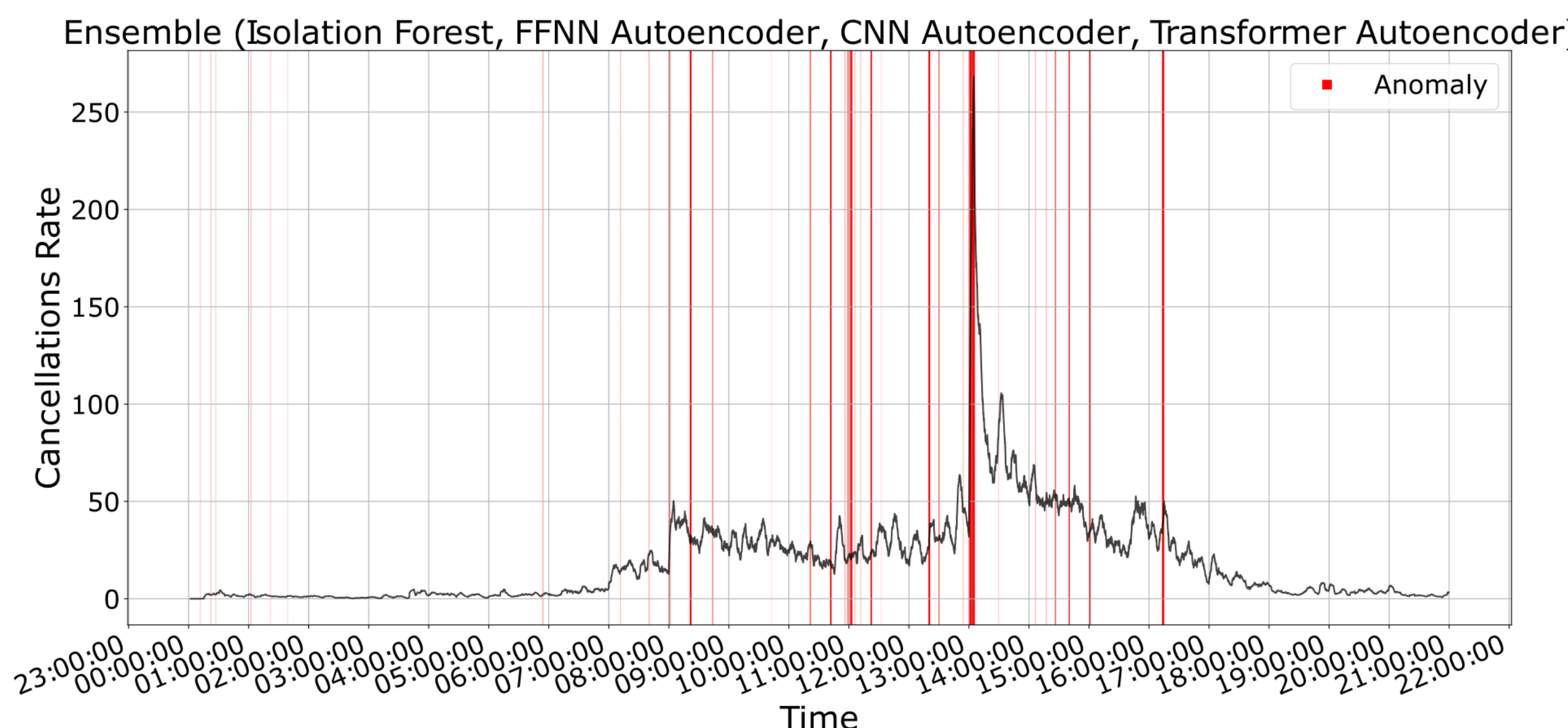
V rámci práce bylo implementováno šest detekčních algoritmů:

- Izolační les (Isolation Forest)
- Lokální faktor odlehlosti (LOF)
- Jednotřídní SVM (One-Class SVM)
- Plně propojený autoenkodér
- Konvoluční autoenkodér
- Transformer autoenkodér

Byla použita reálná historická data z Deutsche Börse AG, která prošla předzpracováním, extrakcí klíčových příznaků a redukcí dimenze.

Modely byly trénovány a hodnoceny pomocí metrik Excess Mass a Mass Volume, které jsou vhodné pro neanotované datové sady. Výpočty probíhaly ve výpočetním prostředí e-INFRA CZ (MetaCentrum).

Součástí řešení je také návrh interaktivního nástroje pro vizualizaci veškerých výsledků.



Detekované anomálie souborem modelů

Úvod

Cílem práce je navrhnout a implementovat rámec pro detekci anomálií v časových řadách z knih limitních objednávek, se zaměřením na odhalení spoofingu – manipulativního chování na finančních trzích. Důraz je kladen na použití metod strojového učení bez učitele nad neanotovanými daty.

Východiska, analytická část

Limitní kniha objednávek (Limit Order Book, LOB) je dynamická databáze, která v reálném čase eviduje příkazy k nákupu a prodeji finančních instrumentů. Obsahuje nabídkovou (Bid) a poptávkovou (Ask) stranu, přičemž objednávky jsou řazeny podle ceny a času zadání.

Spoofing představuje formu tržní manipulace, při níž jsou zadávány rozsáhlé objednávky bez záměru jejich realizace, s cílem vyvolat falešný dojem tržní aktivity. Po dosažení požadovaného cenového pohybu jsou tyto příkazy stornovány.

Vzhledem ke krátkodobému charakteru spoofingu a absenci anotovaných dat je jeho odhalování obtížné. Práce se proto zaměřuje na jeho detekci prostřednictvím metod učení bez učitele.

Dosažené výsledky

Nejlepší výkonnost zaznamenaly modely Izolační les a Transformer autoenkodér. Kombinací nejlepších modelů vznikl robustní systém schopný detekovat potenciálně manipulativní chování bez nutnosti ruční anotace.

Navržený vizualizační nástroj má potenciál pomoci odhalit reálné případy manipulativních praktik a podporuje expertní analýzu a interpretaci detekovaných událostí.

Celý systém je plně škálovatelný a připraven pro nasazení na rozsáhlých datových sadách.

Závěr

Práce představuje funkční řešení pro detekci objednávek podezřelých ze spoofingu pomocí metod strojového učení bez učitele. Aplikace těchto přístupů na finanční časové řady může sloužit jako nástroj pro podporu analytiků a regulačních orgánů při odhalování tržních manipulací.

Do budoucna je možné systém rozšířit o další modely a příznaky vstupních dat. V případě dostupnosti spolehlivě anotovaných dat lze navázat také metodami učení s učitelem.