

Входни данни

Предвид твоя опит като Data Scientist и ML Engineer сме подготвили една задача свързана с числови редици. Прикачвам данни от cash inflow & outflow на дневна база за една от основните сметки на една компания за период от 24 месеца. Има различни валути към тази сметка.

Обхват и цел на задачата

Целта на задачата е да се направи prediction за cash inflow или outflow (по твой избор) за предстоящите 90 дни след края на числовата редица на дневна база. Понеже е тестова задача оставяме на теб да си избереш за коя валута да направиш forecast-a. Няма нужда да бъде правен за всички. По-надолу в документа сме изброили конкретни точки, на които бихме искали да обърнеш внимание.

Критерии за оценяване:

- четимост на кода
- аргументация на подхода
- точност на предвижданията
- описание на стъпките по кода

Указания и въпроси

Първа част

- Анализирайки различните валути, намираш ли сходства и/или разлики между валутите?
- Откриваш ли проблеми с данните, ако да, изброи ги? Какви опции / методи има на групово ниво за решаване на откритите проблеми с данни?
- Групирай времевите редове по избрани от теб критерии и предложи подход за решаването на проблемите/особеностите на групово ниво.
- Дефинирай поне една 'custom' функция и използвай поне един 'loop' при анализа или обработката на данните. (може и във втора част)

Втора част

- Обоснови избора си на валута, за която ще изпълниш задачата.
- Биха ли могли да бъдат класифицирани и ако да по какви критерии?
- Избери метрики и сравни тестваните модели

Резултати / Deliverables

- Очакваме да ни изпратиш кода (модел/и) който е композиран (нотебук) в идеален случай на пайтън
- CSV файл с предвижданията от избора от теб финален модел, който да съдържа колона с деня и със стойността.

Препоръки за следващи стъпки

Как би подходил ако след успешно изработване на модел за една валута получиш зелена светлина да разшириш спектъра на задачата, за да покрива следните характеристики:

- брой валути 15
- входящи и изходящи плащания
- брой компании 10, всяка от които има сметки във над 5 валути до максимум 15 валути

- Оркестриране на множество модели и автоматизиране на техния мониторинг и ъпдейти

Като очакването е предикциите да бъдат доставяни на дневна база за всяка компания/фирма със съответните и валути.