

コーディングテスト (需要予測)

概要

目的

- 機械学習モデルの作成を通じ、プログラミング能力を測る

内容

- 需要予測モデルを構築する

使用言語

- Python(ライブラリは自由にお使いください)

採点ポイント

- コードの熟練度
- 読みやすさ
- 予測精度

提出物

- ソースコード
- 予測結果csvファイル

*予測結果が出せずに終わった場合でも、コードの書き方も評価対象のため、その際はソースコードのみご提出ください。

提出物

提出期限

- 本ファイル受領日から1週間後

提出用ファイル

- ソースコード
 - Pythonファイル(.py)でご提出ください
- 予測結果のcsvファイル
 - ファイルの形式は、sample_submission.csvを参考に作成ください
 - 例：submission.to_csv('my_submission.csv', encoding='utf_8_sig', date_format='%Y-%m', header=False)
 - 下の例のように、提出するCSVファイルには「ヘッダー」が含まれないようにしてください

85列

12行

2020-01	XX	XX	・ ・ ・	XX
2020-02	XX	XX	・ ・ ・	XX
・	・	・	・	・
・	・	・	・	・
・	・	・	・	・
2020-12	XX	XX	XX	XX

対象データ

ある電機会社の2010年1月～2020年12月までの過去販売個数および割引率データを対象にします。
過去販売個数データは月次で記録されており、商品IDや過去の割引率なども確認することができます。
今回の予測ターゲットは2020年1月～12月における84商品における予想販売個数です。以下が使用するデータの概要です。

データの種類と用途

データは①メインデータ、②応募用サンプルファイルの2種類があります。それぞれの用途は以下の通りです。

- ①メインデータ(sales_history.csv、waribiki_history.csv、sales_mikomi.csv、waribiki_mikomi.csv)：2010年1月～2020年12月までの商品・月毎の販売個数及び割引率の実績・見込みデータ
- ②提出用サンプルファイル(sample_submission.csv)：提出時のフォーマットデータ

メインデータの詳細

ファイル	カラム	ヘッダー名称	データ型	説明
sales_history.csv	0	商品ID	int	商品の識別番号
	1	日付	datetime	yyyy-mm形式
	2	販売個数	int	月毎・商品毎の販売個数 (2010/1～2019/10)
sales_mikomi.csv	0	商品ID	int	商品の識別番号
	1	日付	datetime	yyyy-mm形式
	2	販売個数見込み	int	月毎・商品毎の販売見込み個数 (2019/11、2019/12)
waribiki_history.csv	0	商品ID	int	商品の識別番号
	1	日付	datetime	yyyy-mm形式
	2	割引率	int	月毎・商品毎の割引率 (2010/1～2019/10)
waribiki_mikomi.csv	0	商品ID	int	商品の識別番号
	1	日付	datetime	yyyy-mm形式
	2	割引率見込み	int	月毎・商品毎の割引率の見込み (2019/11～2020/12)

※sales_history.csv、waribiki_history.csvについて、一部のデータが欠損している場合があります。

- 1年先までの長期の需要予測となっており、説明変数としては販売個数と割引率のみとなっています。そのため、過去の販売個数実績のトレンドや割引率の影響を考慮してモデルを作成する必要があります
- 販売個数の予測は整数で行ってください
- 評価指標は商品毎に算出したMAPEの平均値で行います（下表参照）

日付	商品ID1	商品ID2	...	商品ID84
2020-01	予測値XX	予測値XX	...	予測値XX
2020-02	予測値XX	予測値XX	...	予測値XX
...
2020-12	予測値XX	予測値XX	...	予測値XX
商品毎のMAPE	MAPE_1	MAPE_2	...	MAPE_84
評価指標	[MAPE_1～MAPE_84]の平均値			