

使用可能なデータや環境と応募用ファイルの説明

本コンペティションで使用できるデータや環境、応募用ファイルのフォーマットについて説明します。

Slack招待リンク

本コンペティションでお知らせや環境提供の際に用いるSlackのリンクは以下になります。

https://join.slack.com/t/gismlit/shared_invite/zt-3ic7gxlsj-sL8q223mBaBFTiSS2gm04A

利用可能データ

配布されるデータおよびその他利用可能データは以下の通りです。

配布データ

コンペティションサイトで取得可能なデータ

Readme

README.pdfとして与えられます。このファイルです。

データ定義書

data_definition.zipとして与えられます。解凍すると**data_definition.xlsx**というファイルが作成されます。LIFULLポータル上に掲載された物件データのカラムについてのデータの定義や説明が記載されています。

学習用データ

train.zipとして与えられます。解凍すると**train.csv**というファイルが作成されます。**train.csv**には、2019年～2022年の1月または7月にLIFULLポータル上に掲載された物件データが含まれています。このデータは、過去の売買価格(**money_room**)や物件の特徴を含んでおり、売買価格予測モデルの開発に使用できます。

評価用データ

test.zipとして与えられます。解凍すると**test.csv**というファイルが作成されます。**test.csv**には、売買価格(**money_room**)を除く2023年の1月または7月にLIFULLポータル上に掲載された物件データが含まれています。**id**というサンプルIDが新たに付与されています。このデータを基に売買価格 (**money_room**) を予測していただきます。

応募用サンプルファイル

sample_submit.csvとして与えられます。予測結果ファイルを作成するときの参考にしてください。

外部データ

コンペティションサイト外で使用可能なデータ

国土数値情報

全国の地理空間情報が含まれるGISデータを以下のサイトより、ダウンロードができます。このデータは、売買価格予測に関連する地域ごとの特徴量を生成する際など広く活用できます。**このコンペティションではこのデータを使用することが必須となるので、ご留意ください。**

国土数値情報ダウンロードサイト : <https://nlftp.mlit.go.jp/ksj/>

登記所備付情報地図データ

登記所備付情報地図データには、不動産登記に関連する地図や図面の電子データが含まれています。このデータは、物件の権利関係や土地の詳細情報を把握するために使用可能です。

登記所備付地図データダウンロードサイト : <https://front.geospatial.jp/moj-chizu-xml-readme/>

提供する分析環境

Snowflake のAIデータクラウド環境

データ共有やAI/機械学習アプリケーションの構築のためSnowflake のAIデータクラウド環境を活用することができます。情報が更新された場合はSlackにてお知らせいたします。

Databricks分析環境

日本語(自然言語)でのデータ解析やダッシュボード作成、高度な分析者向けの予測モデル構築及び推論の実行環境が可能です。情報が更新された場合はSlackにてお知らせいたします。

応募用ファイル

評価対象の(**target_ym**と**unit_id**で指定される)物件の売買価格(**money_room**)を予測し、その結果を以下のような形式で作成してください。

id	money_room
000000	7500000
000001	8000000

idは**test.csv**の**id**と対応しています。これを**ヘッダーなしのCSVファイル**として保存し、コンペティションサイトに提出してください。応募用サンプルファイル(**sample_submit.csv**)も参照してください。

投稿時の注意点

- 応募用ファイルのフォーマットが正しいことを必ず確認してください。
- 投稿後、評価結果が通知されますので、アルゴリズムの改善に役立ててください。