

Amazon SageMaker Canvas ワークショップ

2022/02/09

シニアエバンジェリスト 亀田

[はじめに] Amazon SageMaker Canvas は AWS が提供する構造化データを対象とした機械学習をノーコードで実現できるサービスです。このサービスには初回起動から 2 か月間の無料利用枠が設定されており、本ワークショップシナリオは無料期間内であれば費用は発生しません。ただし SageMaker そのものが用いるストレージの料金は少額が発生しますのでご注意ください。それを過ぎた場合、1 回の学習実行で最低 **30 ドル** の費用が発生することに注意してください。

1. SageMaker のマネージメントコンソールにアクセスします。2022/02/09 現在東京リージョンでは Canvas に対応していないため、バージニア北部、オハイオ、オレゴンなどを選択します
2. 左ペインから Canvas をクリックします

Amazon SageMaker ×

ダッシュボード

検索

Amazon SageMaker Studio

Studio

RStudio

Canvas

3. ユーザープロファイルから[新しいロールを作成]を選びます

新しいロールの作成

カスタム IAM ロールの ARN の入力

既存のロールの使用

AWSNeptuneNotebookRole-0814neptune

AWSNeptuneNotebookRole-0814neptunetest2

AWSNeptuneNotebookRole-20210810neptunenotebook

AWSNeptuneNotebookRole-neptunerolesimple

AmazonSageMaker-ExecutionRole-20220209T112903

AmazonSageMakerServiceCatalogProductsUseRole

redshiftml-EcsTaskExecutionRole-3MB5MPYQ5WAN

redshiftml-Redshift-294963776963-us-east-1

実行ロールを選択してください

⚠ 必須

✓ SageMaker Studio と RStudio の...

✓ SageMaker Studio Project、および...

✓ IAM、または SSO 認証

4. [任意の S3 バケット]を選びます

IAM ロールを作成する

IAM ロールを渡すと、ユーザーに代わって他の AWS のサービスでアクションを実行するアクセス許可が Amazon SageMaker に与えられます。ここでロールを作成すると、[AmazonSageMakerFullAccess](#) 作成する IAM ポリシーで記述されたアクセス許可が付与されます。作成する IAM ロールにより、以下へのアクセスが提供されます。

✓ 指定する S3 バケット - オプション

☐ 任意の S3 バケット
自分のノートブックインスタンスにアクセスできるユーザーが、アカウント内の任意のバケットとそのコンテンツにアクセスすることを許可します。

☒ 特定の S3 バケット
例: `bucket-name-1`、`bucket-name-2`、`i`
カンマ区切り。ARN、`*`、および `/` はサポートされません。

☐ なし

5. [ロールの作成]をおします

6. 以下のように正しくロールが設定されていれば、[送信]をおします

は、AWS で作成できます。

AmazonSageMaker-ExecutionRole-20220209T113167

成功! IAM ロールを作成しました。

[AmazonSageMaker-ExecutionRole-20220209T113167](#)

7. SageMaker Domain の起動中になりますので、起動完了まで待ちます。Domain とは開発環境である Canvas や Jupyter Notebook、EFS ストレージ、IAM ロール、開発環境へアクセスできるユーザーなどが設定としたまとまった SageMaker の設定単位のこと

とを言います。


8. 10 分程度で Domain の設定作業が完了しますので、アプリケーションを起動から [Canvas] を選びます



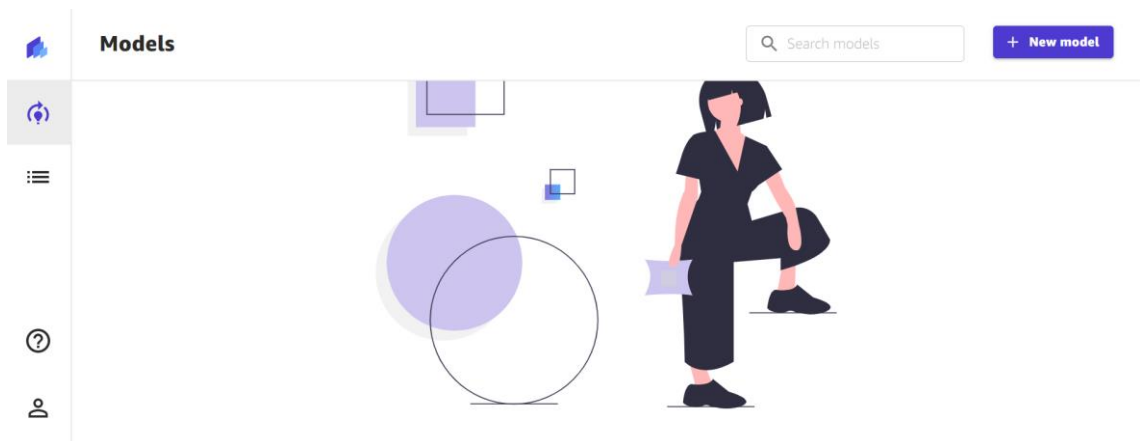
9. 以下のように Canvas の起動画面が表示されますので、起動が完了するまで待ちます。



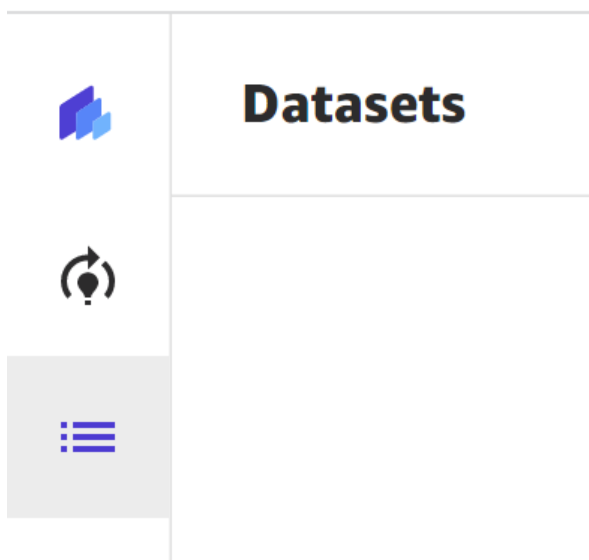
Amazon SageMaker Canvas

Creating the Canvas application default... 

10. 以下の画面が表示されれば起動完了です。



11. 左ペインから [Datasets] をクリックします



12. 画面右上 [Import] ボタンをおします
13. 以下のように Canvas 起動時に自動で S3 バケットが作成されていることがわかります。ブラウザの別タブで S3 マネージメントコンソールにアクセスします

Choose files to import

Amazon S3

Search Amazon S3

<input type="checkbox"/>	Name	Created on ↓
<input type="checkbox"/>	sagemaker-us-west-2-294963776963	02/09/2022 11:40 AM

14. バケット名を特定しクリックします
15. Github から [revised_boston_house_prices.csv] をダウンロードします。このデータは scikit-learn で公開されている米国ボストンの住宅価格を予測するために用いられるデータです。オリジナルデータは、一部差別につながる可能性のある不適切なデータを含んでいるため、その部分を削除しています。

それぞれのカラムは以下を意味しています。

CRIM	人口 1 人当たりの犯罪発生数
ZN	25,000 平方フィート以上の住居区画の占める割合
INDUS	小売業以外の商業が占める面積の割合
CHAS	チャールズ川によるダミー変数 (1: 川の周辺, 0: それ以外)
NOX	NO _x の濃度
RM	住居の平均部屋数
AGE	1940 年より前に建てられた物件の割合
DIS	5 つのボストン市の雇用施設からの距離 (重み付け済)
RAD	環状高速道路へのアクセスしやすさ
TAX	\$10,000 ドルあたりの不動産税率の総計
PTRATIO	町毎の児童と教師の比率
LSTAT	給与の低い職業に従事する人口の割合 (%)

これらのパラメーターを変更した場合、リアルタイムで住宅価格の平均値 (MEDV) を予測するモデルを作ることが今日のハンズオンのゴールです。

16. S3 バケット詳細画面より[アップロードボタン]をおします

sagemaker-us-west-2-294963776963 [Info](#)

オブジェクト | プロパティ | アクセス許可 | メトリクス | 管理 | アクセスポイント

オブジェクト (0)

オブジェクトは、Amazon S3 に保存された基本的なエンティティです。 [Amazon S3 インベントリ](#) を使用して、バケット内のすべてのオブジェクトのリストを取得できます。他のユーザーが自分のオブジェクトにアクセスできるためには、明示的にアクセス権限を付与する必要があります。 [詳細はこちら](#)

17. ファイルの追加、より先程ダウンロードした CSV をアップロードします

ファイルとフォルダ (1 合計, 31.1 KB)

このテーブル内のすべてのファイルとフォルダがアップロードされます。

🔍 名前を検索

< 1 >

<input type="checkbox"/>	名前 ▲	フォルダ ▼	タイプ ▼	サイズ ▼
<input type="checkbox"/>	revised_boston_house_prices.csv	-	application/vnd.ms-excel	31.1 KB

18. 設定などは全てデフォルトのまま[アップロード]ボタンをおします
19. ブラウザ別タブですでに開いている Canvas の画面に戻り、S3 バケットをクリックします
20. 以下のように CSV が認識できていれば、チェックをつけて[Import data]をおします

Choose files to import

Amazon S3 / sagemaker-us-west-2-294963776963 🔍 Search Amazon S3

<input type="checkbox"/>	Name	Last updated ↓
<input type="checkbox"/>	revised_boston_house_prices.csv	02/09/2022 12:10 PM

Choose files to import

Amazon S3 / sagemaker-us-west-2-294963776963 🔍 Search Amazon S3

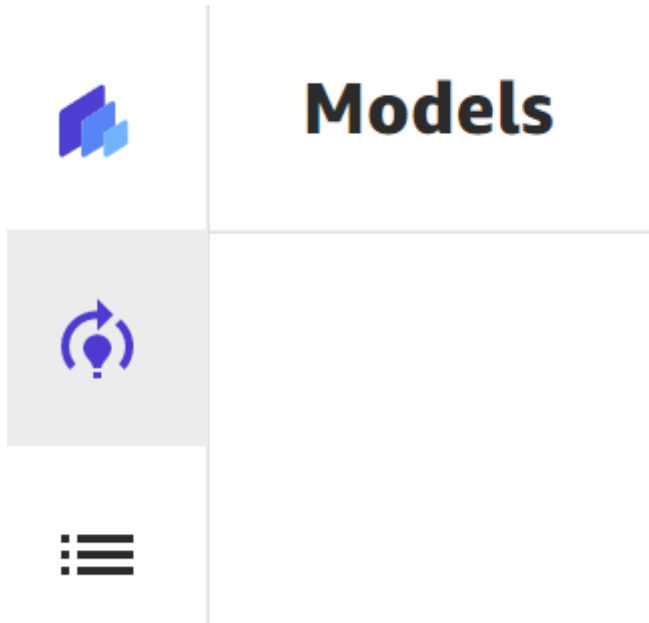
<input checked="" type="checkbox"/>	Name	Last updated ↓
<input checked="" type="checkbox"/>	revised_boston_house_prices.csv	02/09/2022 12:10 PM

1 new dataset [Preview all](#)

21. しばらく待つとデータ取り込みが完了し、Status が[Ready]になれば完了です

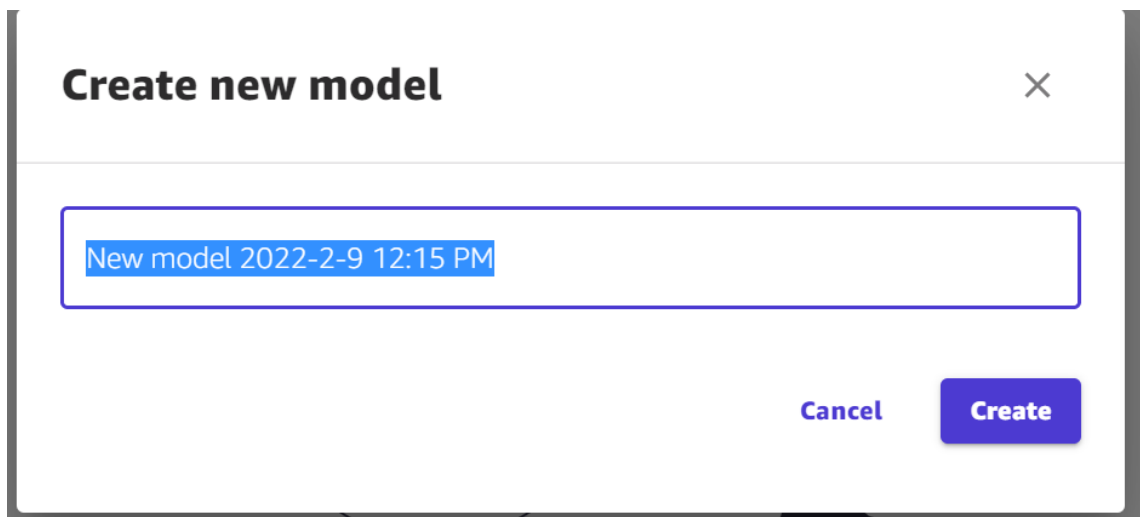
Datasets					
			<input type="text" value="Dataset"/>		Join data + Import
Name	Source	Columns	Rows	Created	Status
revised_boston_house_prices.csv	S3	13	506	02/09/2022 12:12 PM	Ready

22. 次に左ペインから[Model]をクリックします



23. 画面右上[+ New model]をおします

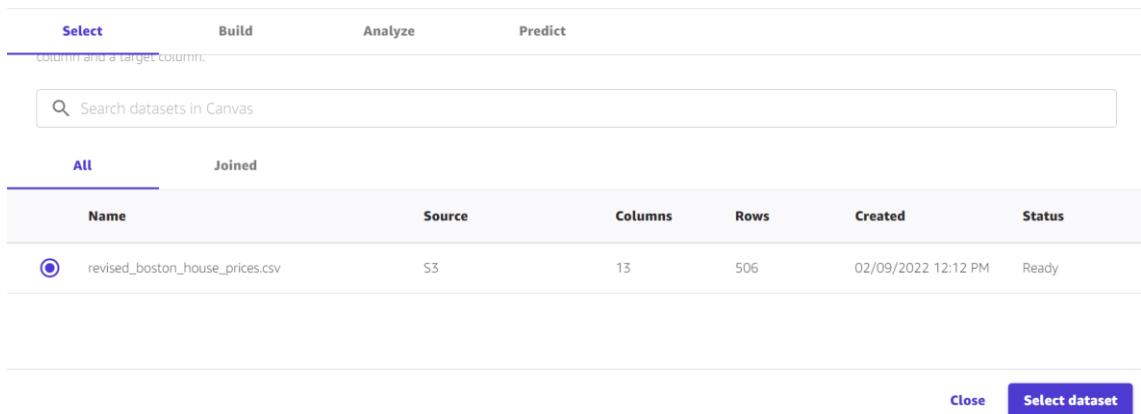
24. 自動で名前が、時間をベースにしていますのでそのまま[Create]をおします



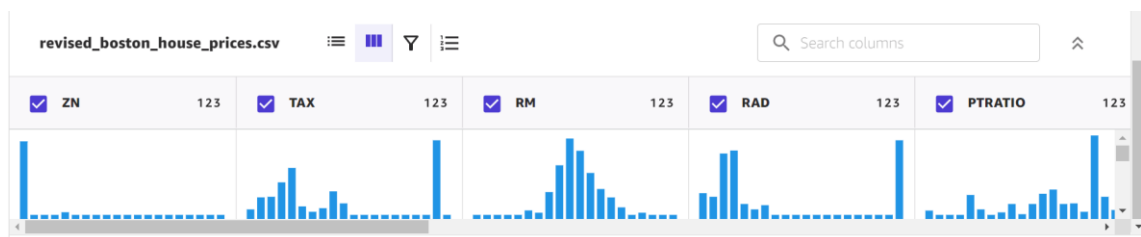
25. Intro ダイアログが表示されますが。Skip します

26. [Select]タブから、先程設定した csv ファイルを選択し、[Select dataset]をおします

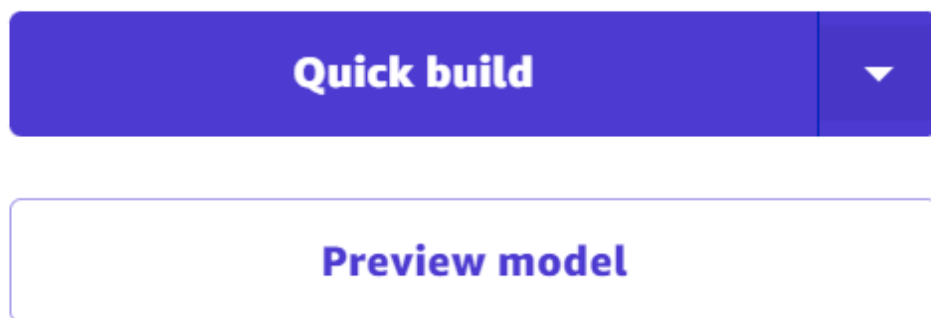
27. Select a column to predict



28. [Build]タブに移動し[Select a column to predict]から[MEDV(住宅価格平均)]を指定します。
29. [Model type]は自動で重回帰分析である Numeric が指定されていますが、[Change Type]を指定すると、2 値分類、多値分類、時系列予測がサポートされていることがわかります。なお、時系列予測は Amazon Forecast を Canvas 経由で起動する形態となっています。
30. 画面下部分では、学習に用いられるデータが偏りを含んでいないか等データ分布状態を確認できるようになっています。



31. [Quick Build]をおします



32. 学習が開始されますので、15 分程度まちます。なおこのハンズオンでは、処理速度を優先させるため Quick Build を選んでいます。商用環境での利用を検討される場合より精度が高くなる Standard Build の指定をおすすめします

New model 2022-2-9 12:15 PM V1 ● Draft + Add version Share ⋮

Select Build **Analyze** Predict

Model overview

Your model is being created. Quick build usually takes 2–15 minutes. You can now leave this view.

Expected build time	Build type	Detailed progress
2–15 minutes	Quick build	Generating column impact

33. 学習が完了すると、以下の通りモデル指標が表示されます。この値は RMSE(Root Mean Squared Error)と言われるもので、回帰分析においては数字が小さければ小さいほど精度が高いとされる指標です。このハンズオンでは 500+件のデータをベースに Quick Build でモデルを作成しているため、それほど精度が出ていないことに注意してください。

Model status

4.7

The model often predicts a value that is within +/- 4.7 of the actual value for MEDV ⓘ

Predict

Share with SageMaker Studio ⓘ

[Advanced Metrics]からより詳細のモデル分析情報を入手可能です。

Advanced metrics

R2 (R-squared) ⓘ	MAE (Mean absolute error) ⓘ	MAPE (Mean absolute percent error) ⓘ	RMSE (Root mean square error) ⓘ
72%	+/-2.94	Not available	4.74

Residuals

Close Download

34. [Predict]ボタンをおします
35. [Single Prediction]をクリックします

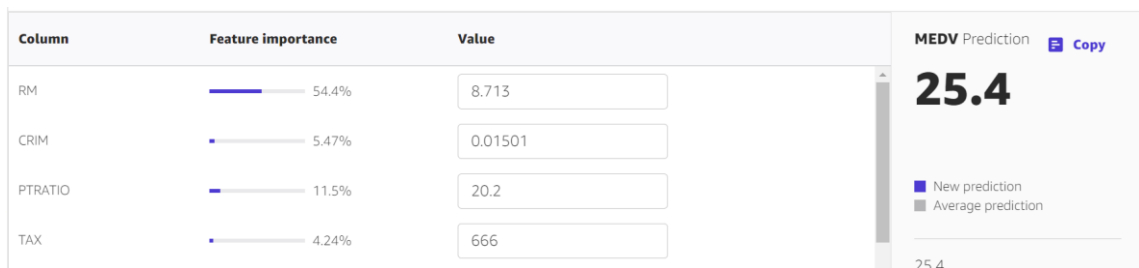
Predict target values

Batch prediction

Single prediction

Modify values to predict MEDV in real time.

36. [Feature importance]とは、予測に影響を与えている各特微量毎の割合を示します。より高い数値のカラムが、より強く予測に影響を与えています。それぞれの値を変更して、画面右[Update]ボタンをおすと MEDV の値が推論されます。



37. (Option)[Batch Prediction]では、複数の予測したい値を csv など準備しアップロードすることで、バッチ予測（複数の値をまとめて予測）することが可能です。例えば、今日の学習に使用した csv をそのまま予測用に指定し実行してみてください。以下のように予測の一覧が表示され、CSV での出力が可能です。

batchInfer-New model 2022-2-9 12:15 PM-revised_boston_house_prices.csv-1644378260								×
Prediction (MEDV)	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	
23.977956771850586	0.00632	18	2.31	0	0.538	6.575	65.2	
20.594505310058594	0.02731	0	7.07	0	0.469	6.421	78.9	
34.6814079284668	0.02729	0	7.07	0	0.469	7.185	61.1	
33.433372497558594	0.03237	0	2.18	0	0.458	6.998	45.8	
36.116432189941406	0.06905	0	2.18	0	0.458	7.147	54.2	
								Download CSV

おつかれしました！

Canvas で作業が終わった際に、必ず左ペイン下に配置されている人型アイコンをクリックして、[ログアウト]をお願いします。ログアウトされずにブラウザを閉じた場合、セッション単位課金が発生し続けます。

削除は以下の順番でおこなってください

- ・ SageMaker Canvas マネージメントコンソールからユーザーをクリック
アプリケーション削除（上記ログアウトを行った場合、自動で削除されます）
 - ・ SageMaker Canvas マネージメントコンソールから Domain を削除
- S3 バケットを削除