

**AutoML 使ってみた**

# ToC

- MLプロジェクトのプロセス
- MLプロジェクトにおける課題
- なぜAutoMLをやるのか
- AutoML の紹介（202008時点）
- 先行事例
- やったこと
- まとめ

# MLプロジェクトのプロセス

- 企画
- 机上検証
- 実証実験
- システム開発
- システム運用

だいたいこんな感じ

# MLプロジェクトにおける課題（の一部）

主に机上検証・実証実験における課題

- データの質・量がよろしくない
- コードの品質が悪い
- 適切なモデルをつくるのが難しい

# なぜAutoMLをやるのか

- データの前処理やクレンジングを担ってくれるのか
- コードの質を気にせずにモデルをつくることができるのか
- 適切なモデルというの確かめることができるか

これらを AutoML が解決できるのかを確かめる。

# GCP AutoML の紹介 (2020/08時点)

- AutoML Vision (画像分析)
- AutoML Video Intelligence (動画分析)
- AutoML Natural Language (自然言語処理) <- 今回はこれやる
- AutoML Translation (翻訳)
- AutoML Tables (構造化データの分析)

# 先行事例

- 画像分類
  - LIFULL, 物件画像の分類
  - cookpad, 商品画像の分類
- その他事例
  - <https://cloud.google.com/automl?hl=ja>
- , etc

## やったこと

- GCP Natural Language AutoML を使う
- データの取得
- データの加工
- AutoMLにデータを投入
- データが読み込まれる



## GCP Natural Language AutoML を使う

- エンティティの抽出
- コンテンツ分類
- 感情分析 <- 今回はこれやる

# データの取得

- 右の画像のようなレビューデータをGASでクロールング
  - Spreadsheet に保存
- 全部で22アプリのレビューデータを取得
- ジャンルは、EC, Game, 人材, 漫画 などなど

## 評価とレビュー

4.1 / 5

1.9万件の評価



★★★★★

よわよわ初心者、2020/05/14

最初からちゃんとやれば楽しい  
自分あんまりこういう系のゲーム長続きしなくてストーリーとか面倒くさくて全部飛ばすしルールも飛ばして感覚でやってました  
でも友達にオススメしてもらって、丁寧にやって [さらに見る](#)

★★★★★

Apple対応最高、2020/05/10

元課金額100万プレイヤー  
かなり昔とは違い、キャラは揃いやすいし天井はあるし新規さんに課金を促しますね。キャラチケで交換出来る欄に、アッシュや樹ミラがいて驚きました。  
今は知りませんが、昔ならどのクエストに入れて [さらに見る](#)

★★★★★

sachi463、2020/07/22

やり込む人向き(レビュー追記しました)  
少し前から始めましたが、今はゴールデンカムイコラボ復刻なので非常に楽しんでいます。  
このゲームはよくあるシンプルなキャラゲーRPGではなく、タイトルに書いたように少しずつ攻略法を把握し [さらに見る](#)

★★★★★

\*氷龍\*、2020/07/15

新規ユーザーです  
幽遊白書コラボをキッカケに始めましたがRPGが嫌いな自分でも久々に面白いゲームだと思いました  
もっと早くに出逢いたかったですw [さらに見る](#)

★★★★★

eidpie、2020/07/04

周回作業が楽しい。  
この手のゲームは続いて半年ですが、グラサマは1年半毎日欠かさずやってます。自分は中課金のほぼソロ専ですが、今はユニットも装備も揃ってきて、クリアできるクエストが増えて成長を感じます。 [さらに見る](#)

★★★★★

風来の食いしん坊、2020/01/13

新規の皆様へ  
皆さんこんにちわ。私はグラサマを初めて1年の者です。課金は2万ちょいしてます、それでは1年プレイして思ったことをつらつらと書いていこうと思います。グラ様をプレイしてて楽しいと感じることはやはり単純な操作では [さらに見る](#)

★★★★★

大奇策-yupi、2020/07/15

楽しんです♡  
僕はグラサマやり始めて1年半になります、これまでプレイした中での感想を述べていきたいと思います...  
このゲームはコラボが多く、新しいクエストや新 [さらに見る](#)

★★★★★

じゅーにとむ、2020/07/14

グラサマ最高(〜♡〜)ニバッ  
グランドサマナーズには3年間もお世話になっています。今でも毎日のように楽しくプレイさせて頂いて、ここまでやり込んだゲームはこれが初めてでした。  
始めは友達に半ば強制的にインストールさせられ [さらに見る](#)

★★★★★

あるりおん、2020/04/11

自分が一番続けたゲーム  
僕は他のゲームのレビューとかしたことないんですが、このグラサマが少しでも人気になるようにかこうと思いましたが！ [さらに見る](#)

## データのラベル付け

- 今回はレビューデータを使うので、あらかじめ星が付いてる。
- したがって、ユーザーが付けた星を正とする。

## ラベルデータがないなら

- 自分でラベルを付ける。
- AutoML Natural Language UI (Data Labeling) などを活用する。
- [AI Platform Data Labeling Service](#) を使用して人間のラベル付け担当者に依頼する。

# データの加工

- 収取したレビューデータをAutoMLが望む形に加工する必要がある。
- いくつかある中で、今回は右のようなフォーマットを選択。
  - X: 1 label - 1 file
  - O: 1 record - 1 file

Source:

[https://github.com/konumaru/sentiment\\_analysis/blob/master/main.py](https://github.com/konumaru/sentiment_analysis/blob/master/main.py)

## ZIP with '0', '1', and '2' scores



# データの投入 (Items)

IMPORT	ITEMS	TRAIN	EVALUATE	TEST & USE	Sentiment Analysis
All items	10,075	<div><div></div>Warning: Importing data</div>			<div>DETAILSDISMISS</div>
Scored	10,065	<div><div></div>Filter table</div>			
Unscored	10				
Training	8,051	<div><div></div>Items</div>			
Validation	1,008	<div><div></div>いまひとつ内容がいまひとつ面白くない。ヤバそうな広告掲載が多すぎる。</div>			
Testing	1,006	<div><div></div>UIが悪い久々に利用しようとしたところ、どのような手続きを踏めばアカウントが復帰できるのかわかりにくい。何度も問い合わせをせねばならず、非常に面倒くさい。</div>			
<div><div></div>Filter labels</div>	<div></div>	<div><div></div>動かない7月25日に不具合が発生しました。起動画面からパスワードも聞かず画面が落ちます。コンビニで使えず、アップデートしてもダメで、財布を取りに戻りました。</div>			
Sentiment score: 0	3,408	<div><div></div>一点だけ不具合iPhone10ユーザーです。CMをみてポイントゲットのコーナーで、音声のみ流れて画像が来ません。運営さんの対応できる点でしたら改善お願いします。あとは文句なしのアプリです。</div>			
Sentiment score: 1	847	<div><div></div>リワードの出方変わったねとてもいいと思います。他のお料理アプリと違い、ほんとに簡単でほんとに美味しい大好きです</div>			
Sentiment score: 2	1,014	<div><div></div>本当にポイント付いてる？カードで購入しても、アプリに加点されない。本当に加点されてる？</div>			
Sentiment score: 3	1,281	<div><div></div>クーポンが役立ちますクーポンをよく利用しています。助かっています。</div>			
Sentiment score: 4	3,515	<div><div></div>使えないチェックインできないし、くじもひけない</div>			
		<div><div></div>定期購読でなければ入れてないとかく広告が多い。無料漫画の最後に広告が出てくるくらいならいいよ。本誌電子版定期購読者なのに、本誌に行くまでも漫画広告を見させられる。やめてほしい。後、すぐキャッシュが膨...</div>			
		<div><div></div>全然繋がらない毎月お金を払っているのに0時に更新しても全然読めない。何とかして欲しい。</div>			
		<div><div></div>いいと思います！クーポンも貰えるし、ニュース速報も入るので、助かります。</div>			
		<div><div></div>EhckcmDnekdc</div>			

# Train

- ワンクリックで実行できる。
- 学習用データ・評価用データをよしなに分割してくれる。
- 学習は何度も実行でき、モデルごとにUnique\_IDが割り当てられる。
- 評価用データにおける精度を確認できる。
- 学習用データの精度がみれないので、過学習の判断ができない。

[START TRAINING](#)

Cloud

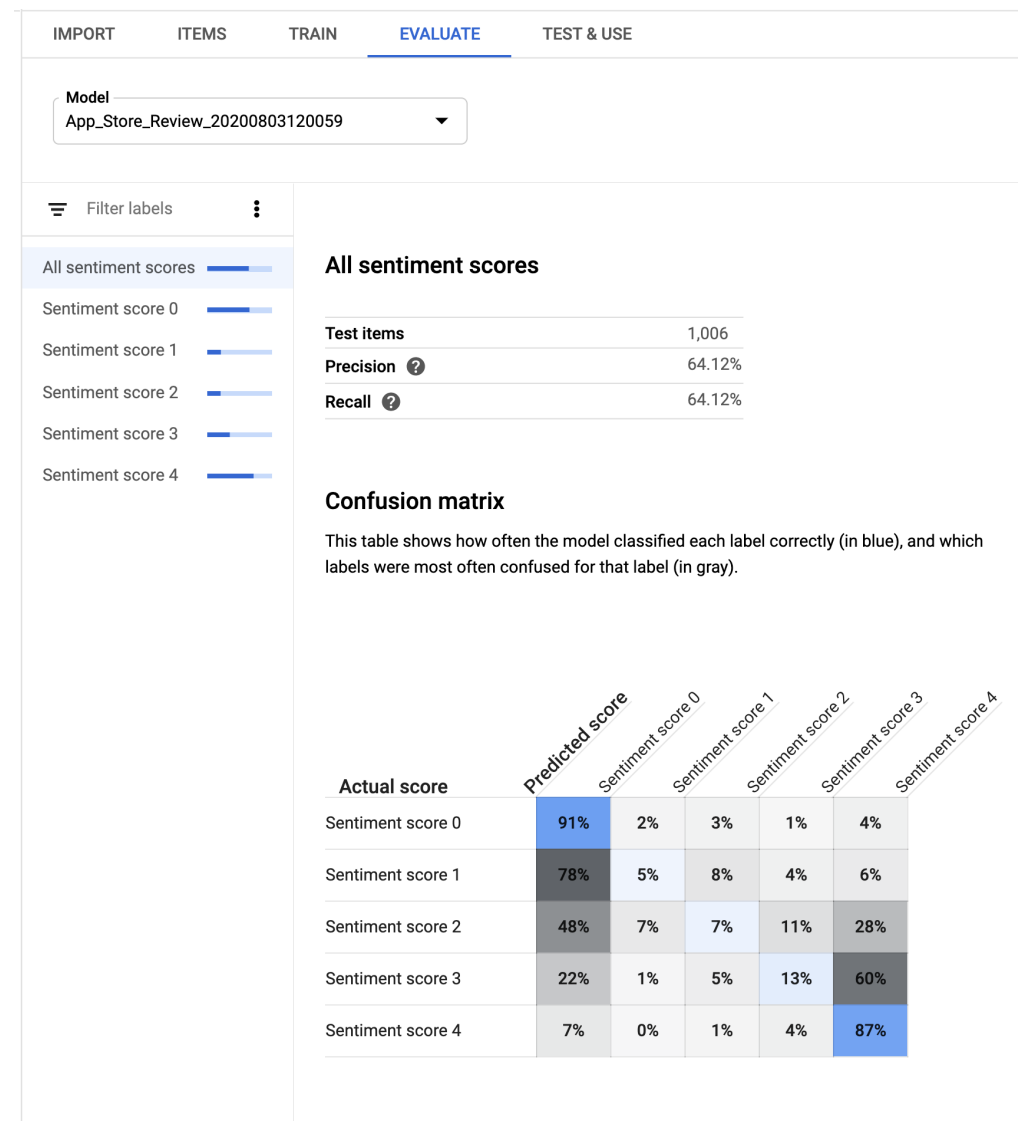


App\_Store\_Review\_20200803120059

**Precision ?**  
64.12%**Recall ?**  
64.12%**Model ID** TST1507124777448374272**Created** Aug 3, 2020, 12:01:05 PM[→ See full evaluation](#)

# Evaluate

- 評価用データの評価結果が見れる。
- 評価用データのラベルごとの数がわかる。
- 評価結果では、Confusion Matrix もみることができる。



# Test & Use

- 簡易的に未知のデータを使って予測することができる。
- 勝手にREST API も作ってくれる。
- 上記を呼べるようなpythonスクリプトの例も出してくれる。

IMPORT ITEMS TRAIN EVALUATE **TEST & USE** Sentiment Analysis

Model  
App\_Store\_Review\_20200803120059

✓ The model was deployed. You can test it now. [REMOVE DEPLOYMENT](#)

**Test your model on new snippets or documents**  
Test your model on text or documents that capture the diversity of your expected inputs.  
[Learn more](#)

☐ Select a file on Cloud Storage

gs:// \* [BROWSE](#)

☒ Input text below

60000 characters remaining

[PREDICT](#)



## Pros / Cons

## Pros

- 決まったデータさえあれば予測モデルからAPIまでつくれる。
- GUIでデータのクレンジングができる。
- GUIで予測結果を探索できる。インタラクティブで楽しい。
- （良し悪しはあるが）学習・評価データを勝手に分割してくれる。

## Cons

- 問題設定が限られるので使い所が難しい
- データの前処理が必要なので no code という訳にはいかない。
- データの前処理が結構たいへん（ここが一番Autoになってほしいな...）
- 自作する場合よりもモデルの解釈が難しい。
- 過学習が判断できない。

# まとめ

- 基本的には便利
- 机上検証・実証実験における課題のうち、コードにまつわる課題は解決されている。

一方で、

- 使い所が限られる。
- モデルの解釈・分析が難しい。
- データの前処理は相変わらず大変。

色々あるが、

今後大いに期待できるツールになるはず。

## おまけ：本当に感情分析するなら

- 感情というあいまいな情報を抽出しようという試みなのでどこかで妥協が必要
  - 機械的な分析を諦め、定性分析を行う
    - > [tensorflow embedding](#)
  - 機械的な分析を諦められないなら
    - > 分散表現を獲得し、k-meansなどでクラスタリングした後、スコアリングするとか？  
(しかし、ものすごくうまく動くことは期待できないと思う)