

DataRobot COVID-19分析官向けプログラム オンライン・セミナー



2020年4月20日

DataRobot Japan

DataRobot

©2020 DataRobot, Inc. - All rights reserved

DataRobot Automated Machine Learning

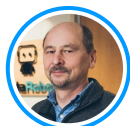
世界最高の機械学習自動化プラットフォーム



Owen Dzung
Product Advisor
Highest: #1
MASTER



Xavier Conort
Chief Data Scientist
Highest: 1st
MASTER



Sergey Yurgenson
Data Scientist
Highest: 1st
MASTER



Amanda Schierz
Data Scientist
Current: #1 Female
MASTER



Jeremy Achin
CEO & Co-Founder
Highest: 20th
MASTER



Tom de Godoy
CTO & Co-Founder
Highest: 20th
MASTER



業界をリードする
データサイエンティストの
経験と知識をソフトウェア化

- **ワンクリック**でデータから機械学習
- 結果を説明、解釈できる **グレーボックス化**
- **デプロイ**までのすべてのプロセスを自動化



COVID-19感染症患者の重症化予測(1)

分析目的

- COVID-19感染症患者さんが将来重症化するリスクを事前に予測するモデルが作れるかどうかを確認する
- 予測できる場合には、どのような因子が予測のために重要なのかを明らかにする

データ

- SIGNATE COVID-19チャレンジで公開されているデータ(2020/4/14ダウンロード分)
- ステータスが明らかになっている437件のデータを対象
- 自由記述欄は除いた

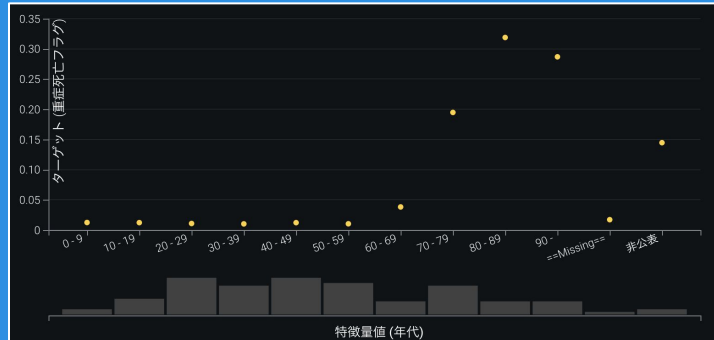
分析方法

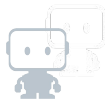
- 予測に意味のない因子やアウトカムとの関係が小さい因子を逐次的に除きながらオートパイロットを合計5回実施

分析結果・考察

- AUCが0.9台(交差検定)のモデルが作成された
- 今回分析したデータでは年齢の重要性が圧倒的に高かった(他には基礎疾患フラグや海外渡航歴など)
- モデル解釈時、例えば平均値を計算するのに使われたサンプル数など統計情報に留意すべき
- 重要性が高いとされた因子のデータを充実させる必要がある

		予測		
		-	+	
実測	-	393 (TN)	8 (FP)	401
	+	15 (FN)	21 (TP)	36
		408	29	437





COVID-19感染症患者の重症化予測(2)

分析目的

- COVID-19感染症患者の症状・経過や行動履歴などの自由記述データから、患者さんが将来重症化するリスクと関係の強い症状が何であることを明らかにする

データ

- SIGNATE COVID-19チャレンジで公開されているデータ(2020/4/14ダウンロード分)
- ステータスが明らかになっている437件のデータ+「症状・経過」や「行動履歴」のテキストを解析し、ステータスを補完してデータを追加(437行->922行)
- 自由記述欄は症状に関連する単語のみを抽出して解析

分析方法

- DataRobotを使ったテキスト解析機能を使ってワードクラウドを作成し、将来重症化に関係の強い単語を確認

分析結果・考察

- 熱や食欲不振、及び咳・呼吸・肺炎など肺に関する症状と重症化リスクには関係性が見られた
- 味覚・嗅覚障害や倦怠感など、一般に言われるCOVID-19感染症の症状と重症化リスクには強い関係性が見られなかった





皆さんは何をしますか？

期待する分析例：感染状況の把握

【基礎統計】

都道府県別、年代別、性別、職業、各種フラグ、濃厚接触者等、またその組み合わせによる各種統計実態（総数や一人当たり）およびトランザクションによる時間的推移を検討。可視化による情報の俯瞰的理解、層別比較や海外データとの比較によるインサイトを抽出。

- ・罹患患者数
 - ・死亡者数
 - ・医療機関受診回数
 - ・検査回数（COVID-19、インフルエンザ、CT、etc）
 - ・自宅待機日数
 - ・利用した公共交通分布
 - ・クラスターサイズ
 - ・体温分布
 - ・症状分布
 - ・発熱観察日、受診日、公表日、回復日、死亡日による期間分析
 - ・海外由来、国内感染の状況
- など

【罹患関係性】

感染者同士や場所、感染経路と拡大など、罹患属性や地理的屬性、時間的推移を検討。可視化による情報の俯瞰的理解、層別比較や海外データとの比較によるインサイトを抽出。

- ・罹患患者ネットワーク
 - ・罹患患者と場所のネットワーク
 - ・ネットワークトポロジー、時間的変化
- など

- まずは分析目的を定義することから
 - 例：「接触からどれくらいで発症するかの予測モデルがあれば、クラスター特定に貢献できる可能性。」
- その分析にはどのようなデータが必要でしょうか？
- SIGNATE様のデータ、別の公開データ、あるいはそれらの組み合わせで！

DataRobot社監修のデータセット

いくつかの公開データをスクレイピングし、クリーニング・構造化

<http://covid-19-public-s3.s3-website-us-east-1.amazonaws.com/>

DataRobot Data Domains COVID-19 Response Efforts

COVID-19 Data Catalog

We have opened up our platform for COVID-19 response effort to provide powerful AI capabilities and accelerate your research for good. Along with this, we are framing public datasources to help enrich your models.

[Learn More >>](#)

ACCELERATING SITUATIONAL AWARENESS [COVID-19]

Datasets & Sources by Domain

Publicly available data can help but needs work. Our DataRobot team has scraped, cleaned, and structured this data. We are providing it to you along with our support.

- COVID-19**
Detailed datasets on COVID infection rates and death rates.
- Citizens**
Public information and census data in support of quality of life.
- Hospitals**
Details on hospitals, beds, and care metric.
- Device**
Granular data from devices and the Internet of Things.
- Reference**
Data to support linking, changing focus, granularity, and cross-referencing.
- Location & Spatial**
All things to support specific location based analysis.
- Public Events**
Calendar events (social distancing, stay at home orders, etc).
- Public Services**
Elements of our public services (transportation, parks & recreation, etc).
- Emergency Services**
Datasets dealing with Police, Fire, and Ambulatory services.



DataRobotホームページのブログ

<https://blog.datarobot.com/jp>

COVID-19に関連したブログを随時公開していきます(現在下記3つ)

- [COVID-19にAI/機械学習で挑む](#)
- [大変動下での機械学習モデルへの対処](#)
- [金融業界のAI活用におけるコロナウイルスの影響](#)



信頼される
エンタープライズ
AIプラットフォーム



DataRobot

信頼される
AIネイティブな
サクセスプログラム