# ブータサイエンス すいすい会

第26回「ロケーションテックと自動機械学習」





# すいすい会の紹介

- ■データサイエンスに関する知見を共有しあい みんなで実践的な理解度を高めたい
- 実践を重視しているGRIなりの考え方を共有します
- 集合知にするため、皆さんの知見をご共有いただきたいです

■資料: GRIホームページ

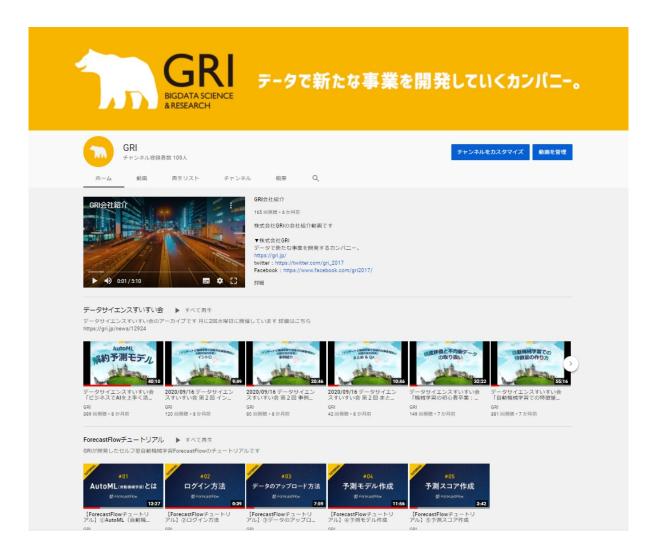
■Slack ForecastFlowチャンネル

https://gri.jp/news/12924

https://join.slack.com/t/forecastflowusers/shared\_invite/enQtNTgyMjcxOTg0NzcxLTBkOWEzYWMwNDJmNTkyMDQzYmIxYWU0YWI4ZmU3ZDU0ZTMxNDUwODAxMWFmYmU1YjJiZGI0MjRhYWYyYTNIZTQ

# 過去のすいすい会の動画(YouTube)

### https://www.youtube.com/c/GRIinc



# ナビゲータの背景

古幡征史 株式会社GRI 取締役



- 株式会社GRI 取締役
- Ph.D in Computer Science
- GRIにて50以上のAI, BI, 分析基盤構築プロジェクトをリード
- KPMGコンサルティング、University of Southern California、ドワンゴを経て、2016年9月より現職

©GRI Inc.

# ブータサイエンス すいすい会

第26回「ロケーションテックと自動機械学習」





## 本日の目的

■ロケーションテックの概観をする

- ■可視化分析との関連性を知る
- ■位置情報を使った自動機械学習の利用例を知る

## ロケーションテックについて

- ■位置情報データの収集技術
- ■位置情報データの蓄積&集計技術
- ■位置情報データの可視化分析技術
- ■地理情報 (GIS: Geographic Info. Sys.)
- ■デジタル地図
- ■位置情報活用アプリケーション(広告など)
- ■O2O/OMO技術

# 位置情報データについて

■人や車などの動きを捉えたデータ(下記は収集方法による分類)

移動体の地点情報を収集

移動体の機器設置場所への近接情報を収集

**GPS** 

GPS

屋外OK 屋内NG 高さNG RFIDタグ



屋内OK 広範囲NG Wifi



屋内OK 広範囲NG ビーコン

(((•)))

屋内OK 広範囲NG

## GPSデータについて

- ■ある時刻での地点データ(緯度、経度)という表現
- ■単一データでは地点にいたのは分かるが、移動と滞在の区別が付かない

デバイスID	データ収集時刻	緯度	経度
HGSJAHASJ	2022-01-16 09:00:00	35.68944	139.69167



## GPSデータについて

- ■複数時刻以上の地点データを組み合わせると滞在・移動の意味を持つ
  - 滞在と移動(移動方法)

デバイスID	データ収集時刻	緯度	経度
HGSJAHASJ	2022-01-16 09:00:00	35.68944	139.69167
HGSJAHASJ	2022-01-16 09:01:00	35.68644	139.69167



# 位置情報データの基礎知識

- ■POI(Point of Interest): 分析対象としての興味のある地点
- ■Geofence: 仮想的な地理的境界線
  - この通りに広告を出した時、どのくらいの人が通過したか知りたくなる
  - ジオフェンスを張って分析をしたい



## GPSデータについて

- ■GPSデータは断続的なため、移動経路までは分からない
- ■POIへの接触を知るためには、ジオフェンスを通過したかの推定アルゴリズムが必要となる
  - GRIでは、このアルゴを自作している(大量データを外部サービスAPIで移動経路推定をかけると、費用がかかりすぎるため)
  - その他に、移動手段推定、利用電鉄線区推定、利用駅推定、利用改札推定



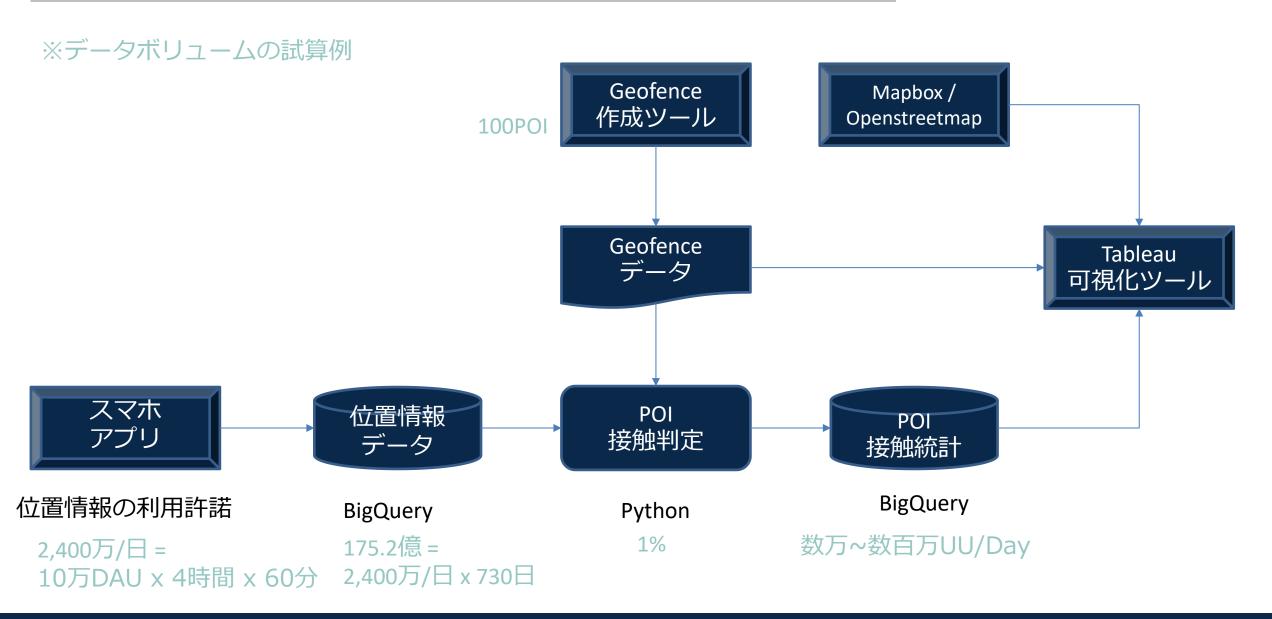
## Tableauでの可視化の例

# ■渋谷駅前のビジョンへの接触統計(ダミー数値)



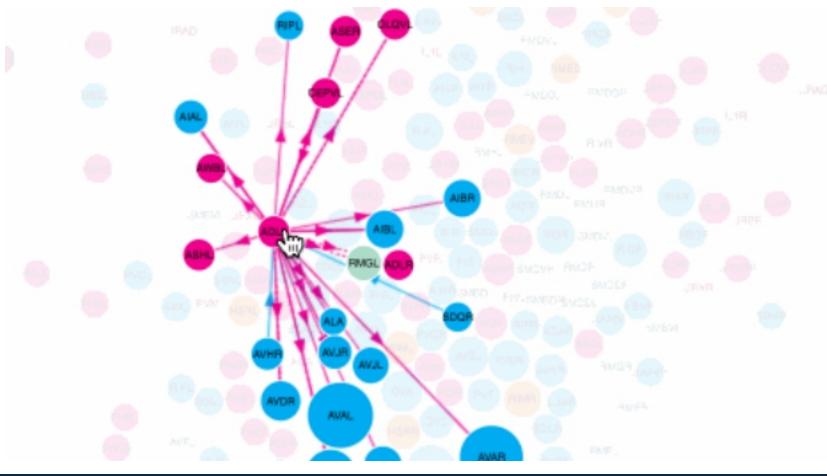
ビデオリサーチ社と共作 https://public.tableau.com/app/profile/griinc6648/viz/\_16424118502200/sheet12

## ジオフェンスへの接触統計レポートのフロー例



# 将来的にグラフDBにて複雑な条件を簡単に集計可視化

- ■Graph DBと可視化で、From-Toのデータ(例:乗車駅-降車駅)を基点と終点で集 約集計、及び可視化できるようにする(専門用語でODペア)
- ■例) 恵比寿駅で降りる人は、どこから来るか?

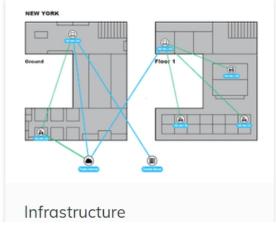


# 将来的にネットワーク分析に強い可視化ツール

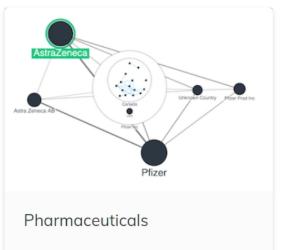
- ■例) Cambridge Intelligence
- https://cambridge-intelligence.com/

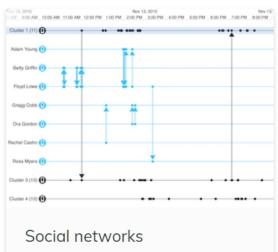










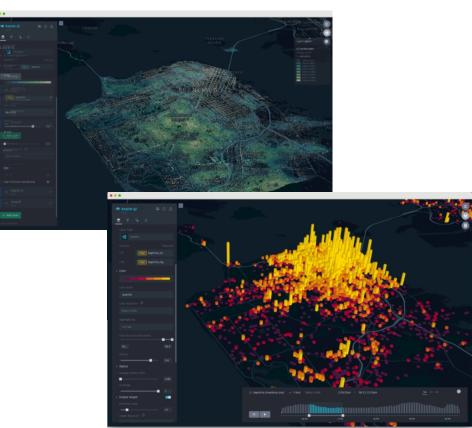


# 位置情報の可視化はjavascriptで発展していくと想定

https://kepler.gl/

- ■Tableauは現在の表形式データ中心のデータの持ち方では限界
- ■UberのKepler.gl (Tableau-kepler.gl拡張も存在)







ロケーションテックと自動機械学習の2つの利用例

©GRI Inc.

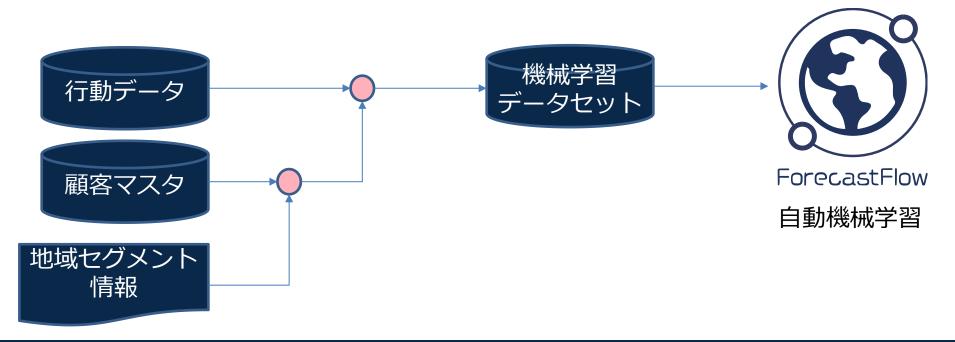


ロケーションテックと自動機械学習の利用例

①顧客の離反予測での地域情報の特徴量利用

# ①顧客の離反予測でのロケーションテックの活用

- ■顧客の住所は、特徴量に入れやすいが、少し効きにくいのと解釈しにくい
  - カテゴリ特徴量の数が47都道府県だと、情報が粗すぎ&読み取りにくい
  - 市区町村や町丁目レベルだと、種類が多すぎるため、読み取れない
- ■顧客住所をGISの地域セグメント情報に変換した特徴量は分かりやすい傾向
  - 技研商事社やMosaic社など(GRIでも構築中)



©GRI Inc.

# 地域セグメントの例

- ■顧客の住所を地域セグメントに変換
- ■把握しやすいのは大きめの粒度で、細かい粒度は把握が困難な傾向

#### ディレクトリーファイル

日本全国約22万件の町丁目(または郵便番号か町字コード)ごとにMosaicのコードが割り振られています。

₹	住所(町丁目)	Mosaic Group	Mosaic Type
1570076	東京都世田谷区岡本一丁目	Α	A01
1570076	東京都世田谷区岡本二丁目	Α	A02
1570076	東京都世田谷区岡本三丁目	В	B03

・・・・Mosaic提供データ3つの種類

① 国勢調査地域版:小地域の町丁字コード

② JGDC版: 国土地理協会の町字コード

③ 郵便番号版:郵便番号コード

А	A01	勝ち組の高学歴 エリート	都会の高層マンションに住むハイクラス・ハイソ サエティな都会暮らしを送るエリート層。
大都市で活躍するエリート 都会の高層マンションに暮らし、日本の経済と文 化を牽引する高学歴・高収入のエリートたち	A02	流行・情報の発信者	カルチャーの発信者として都会に住み、流行・ 情報の最先端を行く、高学歴・高収入な デジタル世代の若者。
В	B03	高級住宅地の エグゼクティブ	豪邸の立ち並ぶ郊外の高級住宅街に住み、裕福な 暮らしぶりの会社役員クラスの核家族世帯。
高級住宅のエグゼクティブ	B04	閑静な住宅街の ファミリー	都心中心部からやや離れた郊外にマイホーム を構える、幼い子供のいるホワイトカラー ファミリー。
大都市郊外の高級住宅地で暮らす大企業で出世 して社会的な地位を手にした裕福な家族世帯	B05	郊外に住む 団塊の世代	郊外に住む、子供が巣立った後の中高年夫婦の ファミリー世帯。
	B06	大都市近郊のホワイト カラーファミリー	首都圏中心部にアクセスのよい住宅地に住む、 若いホワイトカラーの核家族世帯。
	B06	大都市近郊のホワイト カラーファミリー	郊外に住む、子供が巣立った後の中高年夫婦の ファミリー世帯。
С	C07	工業都市の 熟練労働者世帯	大都市圏に住み都心や衛星都市で働く、大きな マイホームを持つブルーカラー世帯。
都市周辺・地方都市の豊かな中高年 都市周辺部にスプロールした住宅地の一戸建て に住む、暮らしぶりの良い回塊の世代の家族世帯	C08	老後の余暇を楽しむ 夫婦	結婚、出産、マイホーム購入、子供の独立、 定年退職というライフステージを順調に重ねた 世帯。
	C09	都市周辺の 兼業農家	住宅地と農地が混在している都市周辺部で農業や 製造業を営んでいる大きな家族世帯。
N	D49	大都市の 公団居住者	UR都市機構によって再開発された大都市圏の公団 住宅に住む人々。
都市部の公営住宅や賃貸アパート住人	D50	長屋住まいの ブルーカラー	大都市の長屋に住む、高齢のブルーカラー世帯。
大都市の公営住宅や長屋に住む、高齢で定収入 なブルーカラーの単身世帯や夫婦世帯	D51	古い子上に団地に住む 高齢者	大都市の公営団地に古くから住むブルーカラー 世帯。
	D52	長屋住まいの 高齢者	大都市の長屋や公営団地に住む、職業の安定しな い中高年の単身世帯。

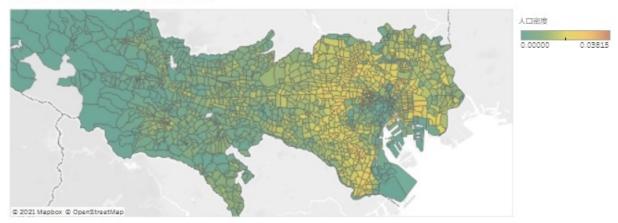
## 郵便番号界とGA界の地図とGISの提供

- ■GRIより郵便番号界とGA界の地図とGISの販売可能
- ■貴社独自データのバンドル販売可能

東京都の人口密度(町丁目ベース)



東京都の人口密度(郵便番号ベース)



eStatの町丁目ベースだと 細かすぎて、分かりずらい (地域形状データを無料で入手可能)

郵便番号単位ならば、 マーケティング的に見やすい粒度

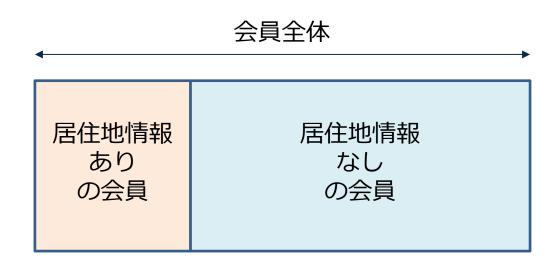
https://public.tableau.com/app/profile/griinc 6648/viz/ 16386259511510/1 1



# ロケーションテックと自動機械学習の利用例

②位置情報から属性推定

- ■スマホアプリの会員から居住地を申告、あるいは配送先として登録してもらっているケースを想定
  - スマホアプリの会員全体の居住県の属性情報があれば、マーケティングや広告配信の規模を拡大することができる
  - スマホアプリで収集できるGPS位置情報から、プライバシーを侵害せず、居住地 を推定する方法のご紹介



- ■居住地情報ありの会員の属性データや行動データで訓練を行い、居住地予測モデルを作成する
- ■居住地予測モデルに居住地情報なしの会員の属性や行動データで居住地の予測スコアを算出する



# ②位置情報から属性推定の例(特徴量)

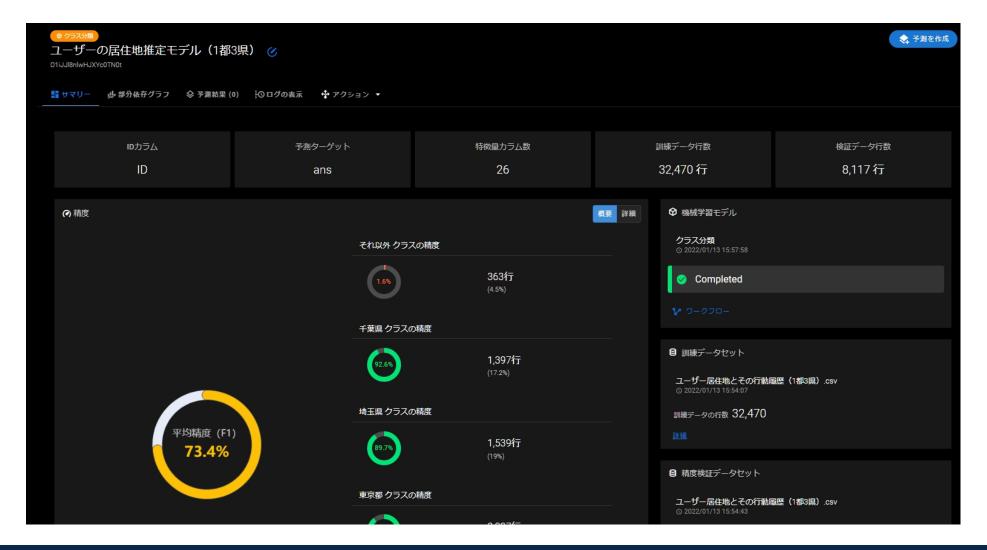
## ■データセット

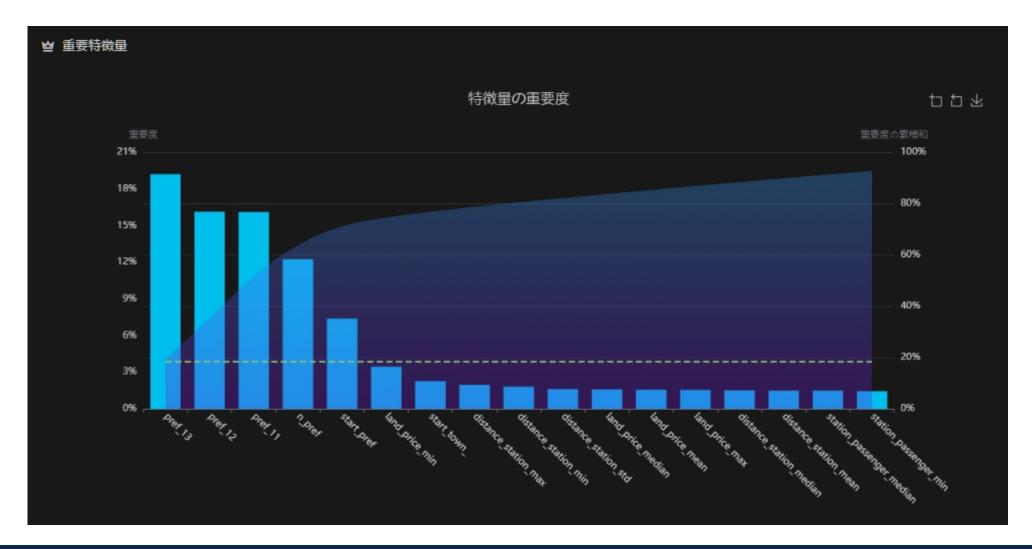
- ID
- ターゲット
  - ユーザの居住都道府県(ここでは1都3県+a)
  - マルチクラス
- 特徴量
  - GPS位置情報をユーザIDごと の統計情報に変換

特徴量	カラム種別	説明 (全てある最新日の)
id	ID	ユーザID
start_pref	特徴量	出発地の都道府県
start_town_	特徴量	出発地の地域
nearest_station	特徴量	最寄駅
n_pref	特徴量	接触判定された都道府県の数
pref_11	特徴量	埼玉県に滞在判定された地点の率(正規化済み)
pref_12	特徴量	千葉県に滞在判定された地点の率(正規化済み)
pref_13	特徴量	東京都に滞在判定された地点の率(正規化済み)
pref_14	特徴量	神奈川県に滞在判定された地点の率(正規化済み)
n_town	特徴量	接触判定された地域(市区町村)の数
n_station	特徴量	接触判定された(半径500m以内)駅の数
is_subway	特徴量	電車を利用したかどうか(n_station >= 3かどうか)
distance_station_max	特徴量	測定された地点とその地点から最も近い駅までの距離の最大値
distance_station_min	特徴量	測定された地点とその地点から最も近い駅までの距離の最小値
distance_station_mean	特徴量	測定された地点とその地点から最も近い駅までの距離の平均値
distance_station_median	特徴量	測定された地点とその地点から最も近い駅までの距離の中央値
distance_station_std	特徴量	測定された地点とその地点から最も近い駅までの距離の標準偏差
station_passenger_max	特徴量	接触判定された(半径500m以内)駅の乗降客数の最大数
station_passenger_min	特徴量	接触判定された(半径500m以内)駅の乗降客数の最小数
station_passenger_mean	特徴量	接触判定された(半径500m以内)駅の乗降客数の平均値
station_passenger_median	特徴量	接触判定された(半径500m以内)駅の乗降客数の中央値
station_passenger_std	特徴量	接触判定された(半径500m以内)駅の乗降客数の標準偏差
land_price_max	特徴量	接触判定された地価の最大値
land_price_min	特徴量	接触判定された地価の最小値
land_price_mean	特徴量	接触判定された地価の平均値
land_price_median	特徴量	接触判定された地価の中央値
land_price_std	特徴量	接触判定された地価の標準偏差
ans	ターゲット	ユーザの居住県

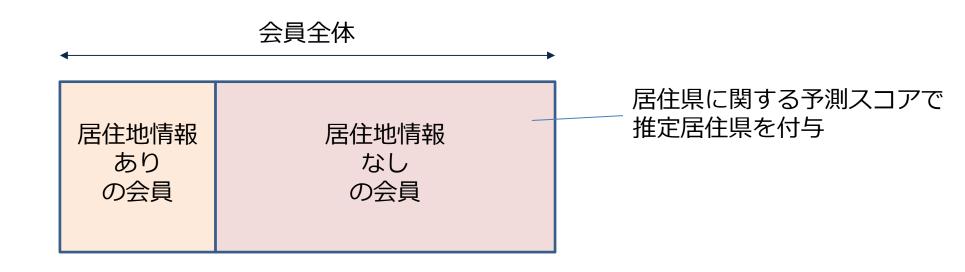
©GRI Inc. 2

全体的に良い精度(その他の地域を除いて)





■居住地情報なしの会員について、機械学習を用いて居住県の推定ができる



©GRI Inc.



# その他の例

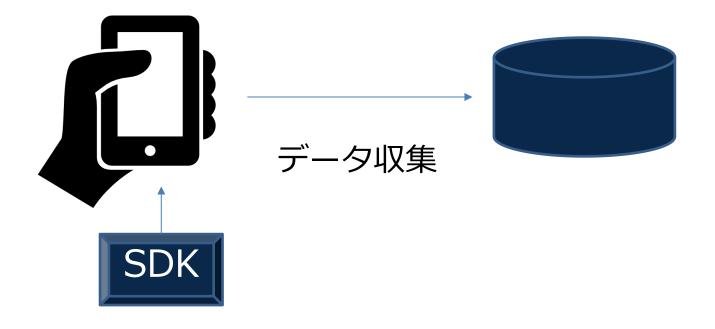
©GRI Inc.

## その他の例

- ■広告による店舗送客の効果分析
  - 結線テレビの情報を使ったCM効果分析を含む
  - キャンペーンによる集客分析
- ■イベントの立地分析
  - 恵比寿と中目黒どちらでイベントを開催すべきか(どのような人が、その街にいるか?)
- ■訪問営業の訪問頻度ごとの営業効果曲線の分析
  - MR、コンサル、保険

# 位置情報データの取得サポート

- 自社アプリにSDKを組み込むサポート
- 位置情報データの販売



©GRI Inc.

## まとめ

■ロケーションテックのパターンの例をご紹介

■現在、パーソナライゼーションの方向性が分析のトレンドだが、これから世帯 やグループでの行動の分析にシフトしていくことが想定される(個人の方が特 定しやすいが、人は社会性のある生き物なので)

■その他の事例など、知りたいことがありましたらお教えください

■また、共有できるお話がありましたら、お教えください

## ForecastFlowの全体的な情報

- ■ForecastFlowの全般的な情報に関しては、下記のホームページを参照してください
  - -https://forecastflow.jp/



## ForecastFlowの無料トライアル

- ■90日間、無料でForecastFlowを試すことができます
- ■下記より、お申込み
- https://forecastflow.jp/

# 次回のすいすい会

- ■2022年2月16日(水)18:30-19:30
- ELTツールMatillionのご紹介
  - Matillionの概要(主にDXを推進するために必要なこと)
  - Matillionの基本機能とデータ分析エコシステムでのMatillionの役割 (SnowflakeやTableauなどとの役割分担)
  - Matillionと自動機械学習ForecastFlowの連携(特徴量エンジニアリングと 予測推論の自動化の話)

# 残っている質問(今後、説明していきたい内容)

- ビジネスの実践例を知りたい
- どのくらい正解データの数があるべきか
- 未来の施策と機械学習の実行タイミングの関係性
- ダウンサンプリングとProbability Calibration
- ForecastFlowのアルゴリズムを知りたい
- 様々な自動機械学習ツールの比較
- 様々なETL/ELTツールの比較
- Feature Store、分析基盤、CDPの説明