

分析提案書

現状と課題	<ul style="list-style-type: none">・ コロナ感染者数の増減によって医療従事者の人員も変化する・ 感染者数が急増した場合、医療従事者の人数が足りていない事態が発生していた・ 事前に感染者数の増減の傾向をある程度予測できれば、あらかじめ人員を用意することができる・ 感染者数の増減は人流が関係していることがわかっている・ また、人流には気象が関係していることもわかっている
データ分析の目的	<ul style="list-style-type: none">・ コロナ感染者数と気象の関係を分析し、感染者数の増減に影響を与える気象要素の推定をする
データ分析の内容	<ul style="list-style-type: none">・ 気象データと東京の感染者数データを結合する・ 散布図や箱ひげ図を描き、気象と感染者数の関係を分析する・ 回帰分析を行い、気象データを用いて感染者数の増減に影響を与える気象要素を予測できるかを検討する
想定される効果	<ul style="list-style-type: none">・ 感染者数の増減に影響を与える気象要素ができれば、天気予報を確認することで感染者数の増減を予測し、人員の準備を行うことで医療従事者の負担を減らすことができる。
添付資料	<ul style="list-style-type: none">・ 添付したノートブックファイル

分析報告書

1. データの確認

分析で明らかにしたかったこと	<ul style="list-style-type: none">気象データ、東京のコロナ感染者数データの欠損値の確認東京のコロナ感染者数の大まかな傾向
分析手法	<ul style="list-style-type: none">棒グラフ、散布図行列
データの種類	<ul style="list-style-type: none">日単位で集計した東京の気象データ日単位で集計した東京の感染者数データ
データの期間	<ul style="list-style-type: none">気象データ：2020年12月27日～2021年12月28日コロナ感染者数データ：2021年1月～2022年12月
補足情報	<ul style="list-style-type: none">分析には東京の気象データとコロナ感染者数データを用いるデータからわかる－新型コロナウイルス感染症情報－ (mhlw.go.jp)

分析報告書

2-1. 月ごとのコロナ感染者数の傾向

分析で明らかにしたかったこと	<ul style="list-style-type: none">月ごとのコロナ感染者数の傾向
分析手法	<ul style="list-style-type: none">クロス集計、棒グラフ、積み立て棒グラフ
データの種類	<ul style="list-style-type: none">月単位で集計した感染者数データ
データの期間	<ul style="list-style-type: none">2021年1月～2022年12月
補足情報	<ul style="list-style-type: none">2022年の感染者数が2021年と比べて極端に多いため、気象データと結合させるときに年度ごとに分割する。

分析報告書

2-2. 曜日ごとのコロナ感染者数の傾向

分析で明らかにしたかったこと	<ul style="list-style-type: none">曜日ごとのコロナ感染者数の傾向
分析手法	<ul style="list-style-type: none">クロス集計、棒グラフ、積み立て棒グラフ
データの種類	<ul style="list-style-type: none">曜日単位で集計した感染者数データ
データの期間	<ul style="list-style-type: none">2021年1月～2022年12月
補足情報	<ul style="list-style-type: none">月曜日の感染者数だけ極端に少ないため、後に月曜日だけ削除する

分析報告書

2-3. 日ごとのコロナ感染者数の傾向

分析で明らかにしたかったこと	<ul style="list-style-type: none">日ごとのコロナ感染者数の傾向
分析手法	<ul style="list-style-type: none">棒グラフ
データの種類	<ul style="list-style-type: none">日単位で集計した感染者数データ
データの期間	<ul style="list-style-type: none">2021年1月～2022年12月
補足情報	<ul style="list-style-type: none">特になし

分析報告書

3-1. 気象データの整理

分析で明らかにしたかったこと	<ul style="list-style-type: none">• 不必要な気象要素の削除• 日付の調整
分析手法	<ul style="list-style-type: none">• 特になし
データの種類	<ul style="list-style-type: none">• 東京の気象データ
データの期間	<ul style="list-style-type: none">• 2020年12月27日～2021年12月28日
補足情報	<ul style="list-style-type: none">• 2020年12月27日～2021年12月26日の気象データは5日、2021年12月29日～2022年28日の気象データは3日分後ろにずらす。

分析報告書

3-2. 日照時間、降水から天気を予測する

分析で明らかにしたかったこと	<ul style="list-style-type: none">日照時間、降水から天気を推定する
分析手法	<ul style="list-style-type: none">クロス集計
データの種類	<ul style="list-style-type: none">東京の気象データ
データの期間	<ul style="list-style-type: none">2020年12月27日～2021年12月28日
補足情報	<ul style="list-style-type: none">降水はすべて雨とする（雪は考慮しない）ゲリラ豪雨のような降水（1時間最大降水量が10mm以上で判断）の日は晴れとした降水がなければ、日照時間で晴れか曇りかを判断する月ごとの最大日照時間とその日の日照時間の比が0.2未満であれば曇り、それ以上は晴れとする

分析報告書

3-3. 感染者数データと結合し集計、可視化する

分析で明らかにしたかったこと	<ul style="list-style-type: none">各気象要素と感染者数の関係
分析手法	<ul style="list-style-type: none">クロス集計、散布図行列、箱ひげ図
データの種類	<ul style="list-style-type: none">東京のコロナ感染者数データ東京の気象データ
データの期間	<ul style="list-style-type: none">気象データ：2020年12月27日～2021年12月28日コロナ感染者数データ：2021年1月～2022年12月
補足情報	<ul style="list-style-type: none">2022年の気象要素と感染者数の増減は無相関であると判断し、2022年のデータは機械学習に用いない

分析報告書

4-1. 特徴選択

分析で明らかにしたかったこと	<ul style="list-style-type: none">• 人流に影響を及ぼす気象要素の特定
分析手法	<ul style="list-style-type: none">• フィルター法、ラッパー法
データの種類	<ul style="list-style-type: none">• 東京の気象データ• 東京の感染者数データ
データの期間	<ul style="list-style-type: none">• 気象データ：2020年12月27日～2021年12月26日• コロナ感染者数データ：2021年1月～2022年12月
補足情報	<ul style="list-style-type: none">• ラッパー法で特徴量が1つしか出なかったため、ラッパー法を採用した単回帰分析とフィルター法とラッパー法を組み合わせた重回帰分析の両方を行う

分析報告書

4-2. 機械学習

分析で明らかにしたかったこと	<ul style="list-style-type: none">分析した特徴量を用いてモデルで精度よく予測できるか
分析手法	<ul style="list-style-type: none">線形回帰（単回帰、重回帰）ランダムフォレスト
データの種類	<ul style="list-style-type: none">東京の気象データ東京の感染者数データ
データの期間	<ul style="list-style-type: none">気象データ：2020年12月27日～2021年12月26日コロナ感染者数データ：2021年1月～2022年12月
補足情報	<ul style="list-style-type: none">予測の可視化には、より精度よく予測できた単回帰分析とランダムフォレストを採用した

考察	<ul style="list-style-type: none">線形回帰、ランダムフォレストともに感染者数を精度よく予測することができなかったことから使用する特徴量が適していないか、東京のコロナ感染者数に与える影響には気象要素以外の要因が大きい可能性があると考えられる。年度によって感染者数が大幅に異なることから、コロナウイルスの性質の違いによって感染者数にも変化が表れていると考えられる。予測精度を上げるためには気象とは別の情報（経済状況など）を分析に用いるか、感染したウイルスの性質によってデータを分けるなど、より多くのデータを分析に用いる必要があると考える。
提案	<ul style="list-style-type: none">どのようなデータがあればより精度よく分析できるかを以下に提案する<ul style="list-style-type: none">日ごとの東京の人流データウイルスの株ごとの感染者数緊急事態宣言や自粛期間中の飲食店の日ごとの売り上げ日ごとのPCR検査の実施状況や実施本数
添付資料	<ul style="list-style-type: none">添付したノートブックファイル