

第 2 回 RapidMiner AI コンテスト 発表会 開催レポート

2024 年 5 月 27 日(月)13 時～、弊社主催の「第 2 回 RapidMiner AI コンテスト 発表会」をオンラインにて開催しました。本レポートは、発表会の内容を要約したものです。

1. はじめに

昨年実施した第 1 回 RapidMiner AI コンテストでは、振動データを用いた車両の分類を実施しましたが、第 2 回 RapidMiner AI コンテストでは、過去の気象予報情報などから一日先の太陽光発電量を予測するモデルを作成し、その予測精度を競って頂きました。2023 年 12 月 20 日から 2024 年 3 月 31 日までの約 3 ヶ月間に渡り実施しました。本発表会では、コンテストで優秀な成績を収めた部門別（20～30 代部門、40 代部門、50 代以上部門）優勝者 3 名の方より分析アプローチについてご報告頂きました。

2. コンテスト概要・結果

本コンテストでの分析課題である「太陽光発電量の予測」という分析課題の出題意図は、需給管理と異常検知の 2 つの実践に活用できるのではないかと考え、分析課題を設定しました。

分析課題の狙いは以下の 3 点です。

- ・天気と風のテキストデータをどのようにテーブルデータに変換するか
- ・日の出、日の入情報からどのように特徴を生成するか
- ・季節の変わり目を予測する課題であり、学習していない範囲をどのようにカバーするか

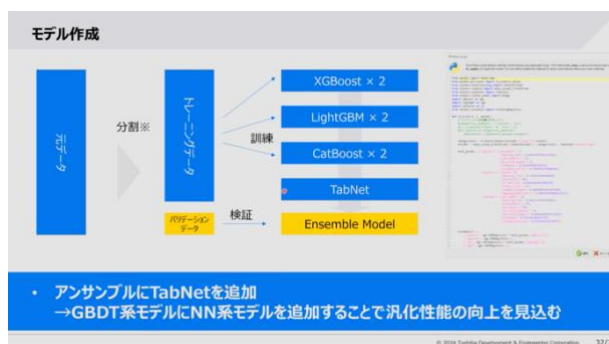
最終的に、約 60 名のエントリーから 28 名の予測結果が投稿されました。実施期間中は最後までスコアが僅差で混戦が繰り広げられ、最高スコアは、東芝デベロップメントエンジニアリング株式会社 國崎様の RMSE : 0.2310 でした。最終順位表は[こちら](#)よりご確認下さい。



3. 発表内容

🏆 20～30 代部門 1 位

國崎 恒成 様（東芝デベロップメントエンジニアリング株式会社 ビジネスデータアナリシスチーム）



■分析アプローチの紹介

太陽光発電量のトレンドと周期性を踏まえ、Window 処理を行い LightGBM による予測モデルを構築されました。データ前処理では、天気予報のテキストデータに対しては ChatGPT を使用して、朝・昼・夜の天気予報置き換え表を作成するなどの工夫もされています（不完全な部分は一部人力で補完）。予測結果を説明変数に加え、それにより発生する誤差の伝播を抑制するために過去の代表的な 1 日の波形を学習させることで周期性を再現しようと試みましたが、曇りや雨の日の精度が低い問題がありました。それら問題を解決するために、日照時間データ（外部データ）を追加することで精度向上につながられたとのことでした（運用時は次の日の日照時間予報を活用）。汎化性能を高めるために GBDT 系モデル¹に加えて NN 系モデルも利用されました。

■感想(発表コメント抜粋)

コンテスト中は、モデルの精度向上に向けて試行錯誤を繰り返しました。特に、代表波形の学習には 2 ヶ月ほどかかり、苦労しました。振り返ってみると、モデルのアンサンブル比率調整やトレンド推定モデルの組み合わせなど、改善の余地はあったと感じています。RapidMiner の使用を通して、プログラムのつながりを視覚的に確認できることの利便性を改めて実感しました。今回の誤差伝播調査では、プロセス全体を俯瞰して考察することができたことが素早い原因特定につながりました。

¹ GBDT は、勾配降下法(Gradient)、ブースティング(Boosting)、そして決定木(Decision Tree)という 3 つの手法を組み合わせた機械学習アルゴリズム。

🏆 40 代部門 1 位

廣井 悟 様（三菱ケミカル株式会社 サプライチェーン所管 オペレーション(日本)本部

九州事業所 企画管理部 生産技術・DX グループ)

今回のプロセス全体概要_1



今回のプロセス全体概要_2



■分析アプローチの紹介

データ前処理については、天気予報と風情報のテキストデータから、独自の書き出し表示のルール設定と時間帯に応じた表現の切り分けを通じて特徴量を抽出されました。特に、前日に発表された天気予報が外れる可能性を考慮し、降水確率の粒度は移動平均を用いて補正されました。季節変化による発電量減衰を考慮するため、日射角度と発電量の関係をモデル化し、変数変換を行うなどの工夫もありました。モデル構築には、Random Forest を用いたモデルを構築され、時間帯による太陽光の日射角度は発電量に大きく影響しており、重要な特徴量として見なされています。降水確率と発電量が大きく乖離する外れ値や外乱を抽出して、それを取り除き再学習を行うことで、モデルの精度向上を実現されました。

■感想(発表コメント抜粋)

自己研鑽と今後の業務への応用を目的にコンテストへ参加しました。RapidMiner は KSK アナリティクスの講習動画で履修していましたが、明確な課題に取り組んだのは初めての経験でした。今回の解析では、Random Forest 以外のモデルのフィッティングに苦労しました。また、色々なアイデアで特徴量や処理をする中で、より効率的なオペレータの活用について課題を感じました。今後は、目的に合った手法を理解し、業務を含めて応用の幅を広げていきたいです。

🏆 50代部門 1位

中谷 香織 様（株式会社 日本触媒 DX 推進部 導入支援グループ）



■分析アプローチの紹介

太陽高度に着目し、時間帯別のモデル構築（Random Forest）を実施されました。太陽高度は、太陽赤緯²や均時差の数表などを用いて計算されたものです。天気予報と風予報のテキストデータに対しては簡単な分類を行い、数値データは扱いやすく変換されました。発電量の実測値と予測値の可視化から、太陽光度が低い時間帯に予測値が下振れしていることが判明し、太陽高度が低い”朝夕”、太陽光度が高い”昼”、太陽光が発電しない”夜”の3つにモデルを分割したところ、精度の向上が見られました。各時間帯のモデルには、Random Forest（朝夕、昼）と Default Model（夜）を採用しました。さらに、データ取得状況を推測して太陽光パネルの設置角を簡易補正しました。前処理サブプロセスの共有化も行い分析の効率化を図られていたことも特徴的でした。

■感想（発表コメント抜粋）

予測結果のずれの原因究明と解決策のアイデア出しに多くの時間を費やしました。特に、太陽光度の計算は労力を要しました。今回の分析では、データの扱いに改善の余地を感じています。今後は、機械学習の理解を深め、様々なモデルやハイパーパラメータの最適化、変数選択を行っていきたいです。

社内にも RapidMiner 愛好者を増やし、次回のコテストでは一緒に参加できるメンバーが増えればもっと楽しいのではないかと考えています。

² 赤緯(Dec, declination)：天の赤道を 0° とし、北極方向を正として±90° まで定めた角度。

4. 総評・コメント

- ChatGPT を用いたデータ処理が印象的で、汎用的な LLM が日常業務にも活用され始めていると感じました。RapidMiner にもその機能が導入されていますので、ぜひお試しください。また、時系列データに特化したモデルに固執せず、問題の本質を理解し通常のテーブルデータとして予測できるモデル選択や変数設計を工夫されていることが印象的でした。(アルテアエンジニアリング株式会社 芝田 隆宏氏)
- LLM 機能はテキストデータ処理を飛躍的に向上させます。RapidMiner の活用事例を紹介する [ブログ](#) でご紹介しておりますのでご参考にしていただければと思います。今後社内でも RapidMiner の利用者とコンテスト参加者を増やしていきたいと考えています。(ネットワンシステムズ株式会社 大河原 昂也氏)
- それぞれ異なるアプローチで課題に取り組んでおられ非常に感銘を受けました。上位入賞者の方には共通点があり、論理的な分析と直感的なひらめきをバランス良く活用して素晴らしい成果を出されていると感じました。発表会は分析の振り返り共有の場として今後も継続していきたいと考えています。次回のコンテストにはぜひ社内の方々と一緒にご参加ください。(株式会社 KSK アナリティクス 森本 好映氏)

5. 最後に

本コンテストにご参加いただいた皆様ならびにご支援いただいた関係各位に心より感謝申し上げます。