

2025 年 11 月 21 日
株 式 会 社 N T T ド コ モ

世界最大級の AI データ分析コンペティション「Kaggle」で金メダルを獲得

株式会社 NTT ドコモ（以下、ドコモ）の鈴木明作（すずきめいさく）、宮木健一郎（みやきけんいちろう）を含むチーム（以下、本チーム）は、世界最大級の AI データ分析コンペティションプラットフォーム「Kaggle（カグル）」^{※1}で 2025 年 7 月から開催された国際コンペティション「Jigsaw - Agile Community Rules Classification」（以下、本コンペティション）^{※2}において、全 2,445 チーム中 11 位に入賞し、金メダルを獲得しました。また鈴木社員は、「Kaggle」コンペティションに参加している約 20 万人^{※3}の上位約 1%に相当する「Kaggle Master^{※4}」の称号を獲得しました。



左から鈴木明作、宮木健一郎

「Kaggle」は、2,700 万人以上^{※3}のデータサイエンティストや機械学習エンジニアなどが登録をしており、実社会の課題に対して最先端のアルゴリズムを活用し、その精度を競う AI データ分析コンペティションプラットフォームです。本コンペティションでは、コミュニティルール配下におけるテキスト分類がテーマであり、オンライン掲示板「Reddit」^{※5}に投稿されたコメントが、コミュニティごとに設けられたルール（広告掲載の禁止など）に違反しているかを判定するタスクが設定され、本チームではコミュニティのルールと投稿文の文脈理解に対して、多様な AI モデルにより予測を行う手法を提案し、タスクの解決につなげることができました。

ドコモでは、ビッグデータを活用した広告ターゲティングのための顧客理解エンジン「docomo Sense[®]^{※6}」など、大規模言語モデルを始めとする AI を活用した技術開発に取り組んでいます。ドコモは、今後もデータサイエンス分野の人材育成に向け、「Kaggle」を始めとするさまざまなコンペティションへの社員参加を推奨し、最先端かつ世界最高峰の AI 技術を習得すると共に、得られた知見や技術を活用した新たな付加価値をお客さまに提供できるよう取り組んでまいります。

※1 <https://www.kaggle.com/>

※2 <https://www.kaggle.com/competitions/jigsaw-agile-community-rules>

※3 2025 年 11 月 21 日時点

※4 Kaggle では個人のメダル獲得枚数に応じて 4 つのランクが設定されており、上から 2 番目の Kaggle Master は金メダル 1 枚と銀メダル 2 枚以上の獲得が条件となっている

※5 <https://www.reddit.com/>

※6 <https://ssw.web.docomo.ne.jp/marketing/strengths/sense/>

*「docomo Sense」は、株式会社 NTT ドコモの登録商標です。

本件に関する報道機関からのお問い合わせ先
株式会社 NTT ドコモ サービスイノベーション部 顧客理解 AI 担当 E-mail : si-customer-analytics-ai-group@ml.nttdocomo.com

本コンペティションの概要

1. テーマ：

本コンペティションでは、オンライン掲示板 (Reddit) のコメントがコミュニティのルールに違反しているかの分類精度を競うテーマが設定されました。参加チームは提供された詳細データから問題設定に適した手法を考案し、そのテキスト分類精度を競いました。

- 提供データ（主なデータを抜粋）

- コミュニティルール：掲示板のコミュニティごとに設けられたルールテキスト（計 6 つ）
 - ✧ 例．広告禁止、法的助言の禁止 など
- 正例サンプル：コミュニティルールに違反しているサンプルテキスト
- 負例サンプル：コミュニティルールに違反していないサンプルテキスト
- 違反判定結果：0 or 1 の二値情報

- 問題設定

本コンペティションでは、以下の図 1 に示す通り、オンライン掲示板である Reddit への投稿コメントが、6 つのコミュニティごとに設けられた各ルールに違反しているかを判定します。

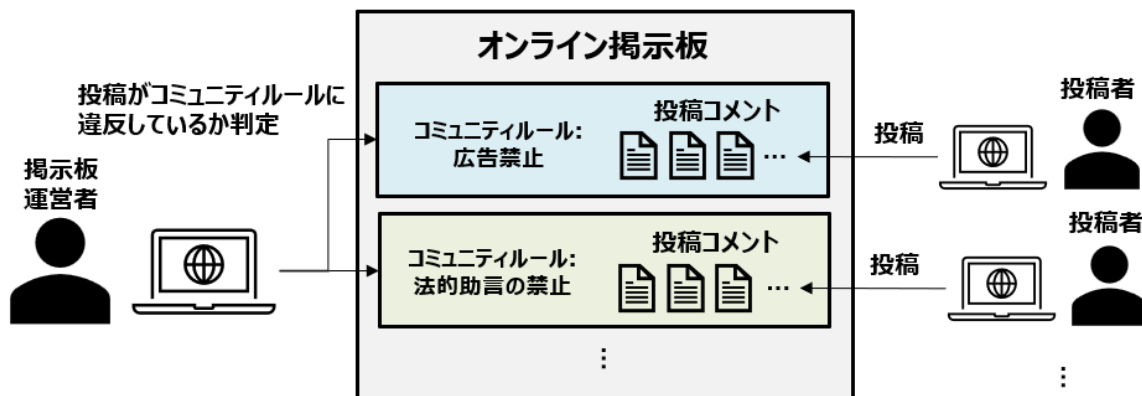


図 1. 本コンペティションの問題設定

2. 提案した解法：

鈴木社員、宮木社員を含むチームは、提供データにあるサンプルが実際に Reddit に投稿された場合に、コミュニティルールの違反になるかという情報を各 AI モデルにオンラインで学習させる手法を元に、今回の解法の検討を行いました。オンライン学習を行う AI モデルには、下記のように得意分野が異なる AI モデルを 13 個使い、各 AI モデルの予測結果を組み合わせ、更に予測対象データを予測後に作成されたラベルを使って再度予測対象データにより学習を行う疑似ラベル (Pseudo Label) や、一部の AI モデルではルールごとに AI モデルを作成するなどの工夫を行うことで、精度が高い予測結果を導く手法を提案しました（図 2）。

- 活用した AI モデル
 - LLM^{※1}：テキストの次の単語を予測することが得意な AI
 - 埋め込みモデル^{※2}：テキストの意味情報の近さを扱うことが得意な AI
 - 分類モデル^{※3}：テキストの分類が得意な AI

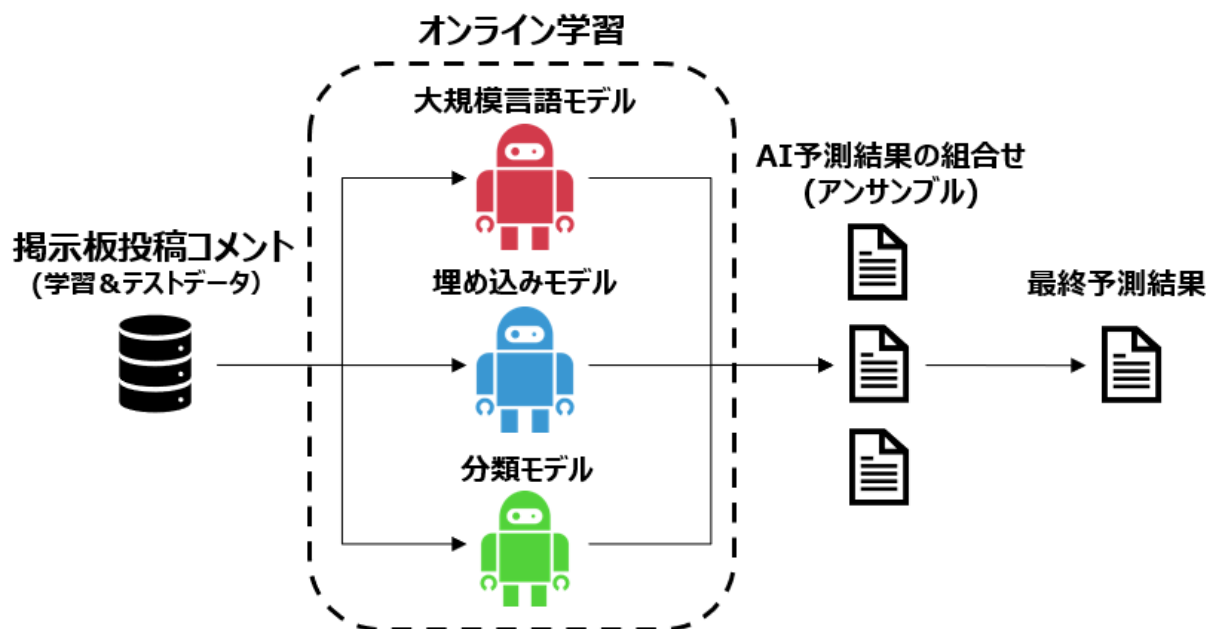


図 2. 金メダル獲得解法

※1 大規模言語モデル

※2 テキストや画像などのデータを数値ベクトルに変換するモデル

※3 入力データをあらかじめ定義されたカテゴリ（クラス）に分類するための機械学習モデル