

## 医薬品個人間取引が疑われる投稿の検出手法の検討

分担研究者 吉田 直子 (金沢大学医薬保健研究域附属 AI ホスピタル・マクロシグナルダイナミクス研究開発センター)

研究協力者 木村 和子 (金沢大学名誉教授 / 一般社団法人医薬品セキュリティ研究会)

張 若愚 (金沢大学大学院医薬保健総合研究科創薬科学専攻)

山口 典枝 (日本アイ・ビー・エム株式会社・グローバル・ビジネス・サービスパートナー)

### 研究要旨

【目的】SNS サイトを介した医薬品等不適正流通の監視手法を開発するため、日本国内で利用されている主な SNS サイトの1つである Twitter (現 X) を対象に、医薬品等個人間取引が疑われる投稿をよりの確に検出するための抽出ロジックを検討した。

【方法】「#お薬もぐもぐ」を検索キーワードに用いたクローリングにより、2023年6月13日時点で2022年1月1日0時0分から2023年3月31日23時59分(日本標準時)までの期間に Twitter 上に投稿され、削除されていない投稿を収集した。収集された各投稿に対し、個人間取引との関係の有無を定義し、テキストマイニングにより文字列を単語ごとに分解して、各単語の出現頻度と医薬品の個人間取引との関連性を調査するとともに、決定木分析の一つである chi-square automatic interaction detection (CHAID) を用いた予測モデルを作成した。

【結果・考察】収集された7,499件のうち、1,709件(22.8%)が医薬品の個人間取引と関係がある投稿であった。テキストを構成する単語の出現頻度を分析した結果、「求」、「譲り」、「手押し」等が医薬品の個人間取引が疑われた投稿に高頻度に用いられていた。また、それらの単語の使用と医薬品の個人間取引が疑われる投稿に有意な関連性が認められた。しかし、個人間取引が疑われる投稿と有意な関連性が認められる単語は多数見つかったため、これらの単語をキーワードとして、医薬品の個人間取引が疑われる投稿を検出することは可能であっても、その作業は煩雑であると考えられた。一方、CHAID を用いた予測モデルの作成を試みたところ、医薬品の個人間取引が疑われる投稿の予測に重要な単語の組み合わせが抽出され、当該投稿において特徴的に使用される単語のパターンが示されたことから、これらを適切に組み合わせたキーワードを用いてクローリングを実行することにより、よりの確に医薬品の個人間取引が疑われる投稿を検出できる可能性が示唆された。

**【結論】**本研究において、医薬品の個人間取引が疑われる投稿の検出手法の一つとして、クローリングプログラムによる投稿情報の収集とその際に使用する検索キーワードの選定方法が考案された。定期的に投稿情報を収集し、直近の情報から抽出された検索キーワードを用いることで、よりの確に監視・指導の対象となる投稿を検出することが可能であると考えられた。

## A. 研究目的

### A-1. 背景

Social networking service (SNS) は、私たちの日常において、インターネットを通じたコミュニケーションを媒介する。これを介して、コミュニケーション以外には、必要な情報の収集や、様々な物品の入手も媒介している。しかし、その便利さゆえに、不適切な投稿なども散見される。これまでに、処方が必要な医薬品、その中には特に向精神薬、そして未承認医薬品などが SNS 上で個人間取引されていたことが確認されている。SNS を介した医薬品、医療機器等の個人間取引は、不適正流通ルートの一つとして認識されており、SNS を介して取引されている医薬品は、正規流通経路を逸脱していることから、適切な保管状態は保障されず、適切な情報提供も得られない可能性があり、当然、低品質医薬品や偽造医薬品の混在リスクも孕むことにより、健康被害をもたらすリスクが高い。医薬品を無資格で譲り渡すことも、薬機法に抵触する可能性がある。国民の保健衛生上の危害未然に防止するため、SNS を通じて行う医薬品取引行為には、積極的に監視指導を行うことが必要である。本研究では、SNS での監視手法の一つとして、医薬品個人間取引に関わる投稿を的確に検出する方法を開発するため、実際の投稿情報のテキストデータを解析し、医薬品の個

人間取引に特徴的なキーワードを抽出し、医薬品の個人間取引に関する投稿を的確に検出するロジックを検討した。

## B. 研究方法

### B-1. 研究対象

日本国内で主に利用されている SNS である Twitter 上の医薬品個人間取引に疑われたテキスト投稿を対象にした。

本調査用に開発したクローリングプログラムを用いて、Twitter 上で医薬品取引が疑われる投稿に高頻度に用いられているハッシュタグ「#お薬もぐもぐ」<sup>1)</sup>をキーワードとして、2023年6月13日時点で2022年1月1日0時00分から2023年3月31日23時59分(日本標準時、以下略)までの期間に投稿され、削除されていない投稿情報を収集した。

### B-2. テキストデータの処理

収集したすべての投稿について、それらの内容や投稿したアカウントに遡って、各投稿において医薬品の個人間取引が疑われる否かを医薬品個人間取引の有無として定義した。

MATLAB Text Analytics Toolbox (MathWorks, Japan) を用いたテキストマイニングにより、投稿されたテキスト情報から単語を抽出し、各単語の出現頻度を集計・解析した。テキストマイニングに

より単語を抽出する際は、カスタムトークン機能により、投稿情報収集時に用いた検索キーワード「#お薬もぐもぐ」を分離しないようにロックした。また、重複4回以下の単語を表示しない設定下で、単語抽出を行った。

### B-3. 単語の出現頻度分析

各投稿の医薬品個人間取引の有無と、各単語の出現の有無とでクロス集計を行い、カイ2乗検定により、有意水準を0.01として、出現単語と個人間取引に関わる投稿との関連性を評価した。解析には、SPSS 19.0.0 (IBM SPSS Inc, Chicago, IL, USA) を使用した。

### B-4. 分類・予測モデルの作成

データマイニングソフトウェア SPSS Modeler version 18.3 (日本アイ・ビー・エム株式会社, 東京, 日本) を用いた決定木分析により、医薬品の個人間取引が疑われる投稿に特徴的に用いられる単語を抽出し、予測・分類モデルの作成を試みた。ホールドアウト検証を目的として、2022年1月1日0時00分から2023年3月31日23時59分(2022年)に投稿された情報から構成されるデータセットを用いて、学習データ:テストデータ=7:3、または、学習データ:テストデータ:検証データ=6:2:2に設定し、決定木分析のひとつであり、カイ2乗検定やF検定に基づいてツリーモデルを構築する chi-square automatic interaction detection (CHAID) により、予測・分類モデルを作成した。作成されたモデルを評価した。さらに、2023年1月1日0時00分から2023年3月31日23時59分(2023年)に投稿された情報

について、作成したモデルによる予測と実際の分類を比較して、判別性能を汎化性能の指標となる受信者動作特性(receiver operating characteristic, ROC) 曲線の曲線下面積(area under the curve, AUC) と決定木による分類の純度を示す Gini 係数(分類前を1として、0に近いほど純度が高い)を用いて確認した。

## C. 結果

### C-1. 収集した投稿情報

「#お薬もぐもぐ」を検索キーワードにクローリングを実行した結果、2022年の1年で収集した投稿数は、7,499件であった。2023年の3ヶ月で収集した投稿数は、6,461件であった。投稿数を分析すると、月別の投稿から、収集時間に近づくほど、収集できる投稿数は多くなる傾向が見られ(Figure 1)、取引と関係有る投稿も収集時点に近づく傾向があることが分かった。

各投稿について、投稿内容が医薬品に関する個人間取引が疑われる否かで定義した結果、2022年の医薬品個人間取引との関係有の投稿数は1,709件(22.8%)、関係無の投稿数は5,790件(77.2%)であった。2023年の医薬品個人間取引との関係有の投稿数は1,625件(25.2%)、関係無の投稿数は4,836件(74.8%)であった医薬品の個人間取引と関係の有無の比に有意な違いは認められなかった( $p=0.1211$ )。

収集した投稿から抽出された単語数は、2022年は3,228個、2023年は2,866個であった。

### C-2. 単語の出現頻度

2022年の投稿情報から抽出された単語3,228個について、クロス集計を行った結

果、各投稿の医薬品個人間取引の有無と各単語の出現の有無について、有意に関連性が認められた単語は 452 個あった (Table 1)。各単語が使用された投稿のうち、個人間取引と関係有の投稿が占める割合で順位付けを行った。高頻度に出現した単語、第 1 位は「求」であり、「求」が使用されていた投稿 83 件中、81 件 (97.6%) が個人間取引と関係有の投稿であった。第 2 位は「譲り」であり、当該単語が使用された投稿総数 103 件のうち、個人間取引と関係有の投稿は 100 件 (97.1%) であった。第 3 位以下は、「手押し」、「頂ける」、「プロフィール」、「譲っ」、「List」、「依頼」、「希望」、「安く」が続き、取引と関係が深い単語として抽出された (Table 1)。なお、クローリング実行時のキーワード「#お薬もぐもぐ」は第 293 位で、当該単語が使用された投稿総数 6,682 件中、個人間取引と関係有の投稿は 1,641 件 (24.6%) であった。

### C-3. 分類・予測モデルの作成

2022 年の全 7,499 件の投稿を目的変数として「医薬品個人間取引疑いあり : 1」(n=1,709) と「医薬品個人間取引疑いなし (n=5,790) : 0」(n=5,790) に定義して、学習 : テスト : 検証データの割合を調整して、CHAID を用いた医薬品個人間取引に関する投稿の予測・分類モデルを作成した。その結果、医薬品の個人間取引に関係のある投稿を特徴づける単語の組み合わせが抽出され、CHAID による予測モデルが作成された。

学習データ : テストデータ = 7 : 3 で生成したモデルは、予測変数として重要度の高い方から順に、「薬」、「DM」、「#メン

ヘラ女子」、「RT」、「#病み垢さんと繋がりたい」、「精神」、「#お薬もぐもぐ」、「リンク」、「@yakubutsumhlw」が抽出された (Figure 2)。「医薬品個人間取引関係有」の投稿に使用される単語の特徴の 1 パターンとして、「DM」の記載がある投稿における医薬品個人間取引率 (1208/5231 件, 23.1%) が最も重要性の高い単語として抽出された (Figure 3)。そのモデルの精度を確認したところ、学習データで正解率は 80.4%、テストデータで 82.3% の正解率が得られた (Table 2)。構築したモデルが実際に医薬品の個人間取引に関わる投稿か否かを分類できるかの汎化性能の評価として、学習データとテストデータにおける AUC は、それぞれ 0.83 と 0.84 程度であり、Gini 係数は、それぞれ 0.65 と 0.68 であった (Table 3)。

学習データ : テストデータ : 検証データ = 6 : 2 : 2 に設定した場合、得られた決定木モデルは、予測変数として重要度の高い方から順に、「薬」、「DM」、「#病み垢さんと繋がりたい」、「#メンヘラ女子」、「RT」、「ください」、「下さい」、「精神」、「@yakubutsumhlw」、「リンク」が抽出された (Figure 4)。「医薬品個人間取引関係有」の投稿に使用される単語の特徴の 1 パターンとして、「DM」の記載がある投稿における医薬品個人間取引率 (1045/4452 件, 23.5%) が最も重要な単語として抽出された (Figure 5)。作成され予測モデルの精度を確認したところ、学習データの正解率は 80.1%、テストデータで 83.8%、検証データは 80.8% の正解率が得られた。構築したモデルが実際に医薬品の個人間取引に関わる投稿か否かを分類できるかの汎化性能の評価として、学習データ、テストデ

ータおよび検証データの AUC は、それぞれ 0.83、0.84 および 0.83 程度、Gini 係数は、それぞれ 0.65、0.68 および 0.67 であった。

作成された予測モデルで正しく予測できるか確認するため、2023 年のデータセットを用いて、予測モデルの検証を行った結果、6,356 件 (84.8%) の投稿は定義と一致する予測結果となり、158 件 (2.1%) は、個人間取引と関係無と定義された投稿が関係有と誤って予測され、985 件 (13.1%) は、個人間取引と関係有と定義された投稿が関係無と誤って予測された (Figure 6)。

#### D. 考察

本研究では、医薬品個人間取引に関する投稿を的確に検出できるロジックを検討するため、「#お薬もぐもぐ」を検索キーワードとして抽出される投稿情報を対象に、当該投稿に使用される単語とその出現頻度、ならびに各単語と医薬品個人間取引に関する投稿との関連性を統計解析や機械学習により解析し、医薬品個人間取引が疑われる投稿を検出する手法の精度向上に向けて、検討を行った。

クローリングプログラムを用いて収集した投稿に使用されていた単語のうち、個人間取引が疑われる投稿と有意な関連性が認められる単語は多数見つかった (Table 1)。これらの単語をキーワードとして、医薬品の個人間取引が疑われる投稿を検出することは可能であっても、単語数が膨大で、その作業は煩雑であると考えられた。個人間取引に有意に関連する多数の単語を特徴付ける単語をさらに絞り込む必要があると考えられた。

一方、機械学習を活用することにより、

膨大なデータから、検出すべき投稿を特徴付ける単語を絞り込むことができた

(Figure 2, 4)。監視手法の開発という目的においては、関連性が高いキーワードが少数に絞られて抽出されるメリットは大きく、特徴的に出現する単語のパターンが示された CHAID による機械学習の方が、出現頻度から統計解析により特徴的な単語を抽出するより、優れていると考えられた (Figure 3, 5)。

本研究で CHAID により予測モデルを作成するにあたり、学習データ：テストデータ=7:3 の結果と学習データ：テストデータ：検証データ=6:2:2 の結果は、どちらも 80% 以上の正解率に達し (Table 2)、同じく「DM」が医薬品の取引に関係のある投稿に特徴的に用いられている単語の 1 つとして、ツリーの第 1 層に抽出された。学習データ：テストデータ：検証データ=6:2:2 は、「DM」に続く第 2、3 層で「ください」や「下さい」が抽出されたが、投実際の稿では、「ください」は主に「DM」の後に記載されていることを考慮すると、学習データ：テストデータ=7:3 と、学習データ：テストデータ：検証データ=6:2:2 で作成された予測モデルにおける差異は小さいと考えられた。また、2023 年の 3 か月分の投稿情報で検証を行う計画であったため、今回は 2022 年のデータセットでは、機械学習で一般的な学習データ：テストデータ=7:3 を使用してモデルを作成した。モデルの評価指標とした AUC は 1 に近づくほど、完全予測モデルとなり、正確に判別できていることを示し、今回作成されたモデルは、取引と関係有無の判別に有用であると考えられた (Table 3)。しかし、

本研究で得られた予測モデルでは、取引と関係有る投稿が関係無と予測される誤りが13.1%生じた (Figure 6)。今回得られたモデルはある程度予測はできたが、精度向上の余地がある。医薬品の個人間取引の監視においては、取引に関係する投稿を見落とさない検出法の開発が求められる。

予測分類モデルの作成の過程において、重要度の高い予測変数の1つとして抽出された「@yakubutsumhlw」は、厚生労働省医薬・生活衛生局監視指導麻薬対策課に委託を受けて(一社)偽造医薬品等情報センターが運営するあやしいヤクブツ連絡ネットのアカウントであり、SNSにおいて、不適正な医薬品取引等に対し、注意喚起を目的とした投稿を行っている (Figure 2, 4)。当該投稿に医薬品のSNS上での取引に関わる記載が含まれるため、本研究におけるクローリングにおいて、投稿情報が収集されたと考えられる。「@yakubutsumhlw」は、医薬品の個人間取引に関係しない投稿に登場する単語として抽出されており、個人間取引に関わる投稿を検出する際には、除外検索 (NOT 検索) の候補となる (Figure 3, 5)。

抽出されたキーワードが的確に医薬品の個人間取引が疑われる投稿を検索できるかどうかを検証するために、2023年3月以降に新しい投稿の収集を試みた。しかし、クローリングプログラムを実行ができない状況となっており、その原因として、TwitterがXに変更となり、APIなども変更されたと考えられた。また、ハンドサーチで「#お薬もぐもぐ」を検索したところ、2023年以前の投稿が大量に削除処理された状況となっており、収集でき

る投稿数は非常に少なく、作成したモデルの検証を行うことは困難であった。SNS運営側が、不適切な投稿やそのアカウントに対し、何等かの対応をした可能性も考えられた。

本研究において、機械学習を活用して、Twitter上の投稿において医薬品の個人間取引が疑われるか否かを予測するにあたり、重要度の高い単語の組み合わせが抽出され、医薬品の個人間取引が疑われる投稿において特徴的に使用される単語の一つのパターンが示された。また、当該投稿に特徴的に使用されない単語も抽出されたことから、これらを適切に組み合わせたキーワードを用いてクローリングを実行することにより、よりの確に医薬品の個人間取引が疑われる投稿を検出できる可能性が示唆された。直近になるほど、収集された投稿数が多かったことから (Figure 1)、今後、最新の投稿情報を収集して、データセット更新し、再学習させる、また、学習データの抽出条件等を検討することで、この誤りがより生じないモデルの構築が実現できると考えられる。また、得たい情報、解析の目的に合わせて、最適な方法を選び、より精度の高い予測モデルの作成、実装につなげることが重要であると考えられる。

本研究の限界として、本研究の対象として、2022年に投稿された情報を収集しているが、クローリング実行時に削除されずに残っている投稿だけが収集されており、投稿後削除された情報は含まれていない。また、投稿情報を収集するプロセスにおいて、Twitter上のAPI制限により、クローリングプログラムで複製・保存できるテキスト量が限られており、長文の

投稿については、すべてのテキストデータを収集できていない。そのため、検索キーワードである「#お薬もぐもぐ」の出現頻度が、収集した投稿数を下回っている。同様の事象が他の単語においても生じていると考えられる。また、出現回数4回以下の単語は、フィールド数が過多になるため、解析対象から除外した。その中に医薬品個人間取引に強く関連する単語があった可能性は否定できず、他のアルゴリズムを活用することで、より多くの情報を網羅した高精度な予測モデルを作成できる可能性はあると考える。しかし、本研究において、検出目的とする投稿を予測するモデルを作成できたことから、出現回数4回以下で制限したことは不適當ではなかったと考える。

## E. 結論

本研究において、医薬品の個人間取引が疑われる投稿の検出手法の一つとして、クローリングプログラムによる投稿情報の収集とその際に使用する検索キーワードの選定方法が考案された。定期的に投稿情報を収集し、直近の情報から抽出された検索キーワードを用いることで、よりの確に監視・指導の対象となる投稿を検出することが可能であると考えられた。

## F. 研究発表

### 1.論文発表

なし

### 2.学会発表

なし

## G. 参考文献

前川京子, Rahman Sofiqur, 朱姝, 木村和

子, 吉田直子, 張若愚, 森本剛: SNS サイトを介した医薬品の個人間取引実態調査. 厚生労働科学研究費補助金 医薬品・医療機器等レギュラトリーサイエンス政策研究事業「医薬品等のインターネット販売に対する監視手法の研究」研究代表者 吉田直子, 令和4年度総括・分担研究報告書, p. 47-72, 2023年3月31日.

Table 1. 2022 年データセットから抽出された単語のうち、医薬品個人間取引と関係有の投稿において、有意に出現率が高かった単語

(計 452 個、並びは取引と関係有の投稿中当該単語の出現した投稿数/出現した総投稿数の割合の高い順)

順位	抽出された単語	出現した総投稿数 (件)	取引と関係有の投稿中当該単語の出現した投稿数 (件)	取引と関係無の投稿中当該単語の出現した投稿数 (件)	取引と関係有の投稿中当該単語の出現した投稿数/出現した総投稿数の割合	p 値
1	求	83	81	2	97.6%	<0.01
2	譲り	103	100	3	97.1%	<0.01
3	手押し	28	27	1	96.4%	<0.01
4	頂ける	25	24	1	96.0%	<0.01
5	プロフィール	23	22	1	95.7%	<0.01
6	譲っ	131	125	6	95.4%	<0.01
7	List	42	40	2	95.2%	<0.01
8	恒例	21	20	1	95.2%	<0.01
9	啓蒙	20	19	1	95.0%	<0.01
10	依頼	19	18	1	94.7%	<0.01
11	各種	30	28	2	93.3%	<0.01
12	希望	53	49	4	92.5%	<0.01
13	ゾル	12	11	1	91.7%	<0.01
14	参照	23	21	2	91.3%	<0.01
15	アンケート	11	10	1	90.9%	<0.01
16	安く	70	63	7	90.0%	<0.01
17	〈アカウント名〉	40	36	4	90.0%	<0.01
18	金額	20	18	2	90.0%	<0.01
19	いただける	20	18	2	90.0%	<0.01
20	郵便	10	9	1	90.0%	<0.01



21	025	10	9	1	90.0%	<0.01
22	ツイプロフ	10	9	1	90.0%	<0.01
23	URL	39	35	4	89.7%	<0.01
24	条件	29	26	3	89.7%	<0.01
25	dm	19	17	2	89.5%	<0.01
26	5st	18	16	2	88.9%	<0.01
27	デエビコ	18	16	2	88.9%	<0.01
28	下さる	9	8	1	88.9%	<0.01
29	別名	9	8	1	88.9%	<0.01
30	レンボレキサント	9	8	1	88.9%	<0.01
31	バチバチダウン	9	8	1	88.9%	<0.01
32	局	9	8	1	88.9%	<0.01
33	凍り	9	8	1	88.9%	<0.01
34	っぱなし	9	8	1	88.9%	<0.01
35	下っ	9	8	1	88.9%	<0.01
36	d	9	8	1	88.9%	<0.01
37	s10	9	8	1	88.9%	<0.01
38	応募	9	8	1	88.9%	<0.01
39	探し	74	65	9	87.8%	<0.01
40	固	40	35	5	87.5%	<0.01
41	ぱ	16	14	2	87.5%	<0.01
42	2m	8	7	1	87.5%	<0.01
43	割引	8	7	1	87.5%	<0.01
44	いくら	8	7	1	87.5%	<0.01
45	申し訳	8	7	1	87.5%	<0.01
46	手渡し	8	7	1	87.5%	<0.01
47	買わ	8	7	1	87.5%	<0.01
48	匿名	8	7	1	87.5%	<0.01
49	〈アカウント名〉	8	7	1	87.5%	<0.01
50	≠切	8	7	1	87.5%	<0.01
51	〈アカウント名〉	8	7	1	87.5%	<0.01
52	#UNCH	8	7	1	87.5%	<0.01
53	求め	39	34	5	87.2%	<0.01
54	DM	381	332	49	87.1%	<0.01

55	#ルネスタ	30	26	4	86.7%	<0.01
56	格安	15	13	2	86.7%	<0.01
57	#コントミン	14	12	2	85.7%	<0.01
58	125	7	6	1	85.7%	<0.01
59	リ	7	6	1	85.7%	<0.01
60	次第	7	6	1	85.7%	<0.01
61	〈アカウント名〉	7	6	1	85.7%	<0.01
62	更新	48	41	7	85.4%	<0.01
63	使え	27	23	4	85.2%	<0.01
64	交換	20	17	3	85.0%	<0.01
65	需要	20	17	3	85.0%	<0.01
66	#エチゾラム	32	27	5	84.4%	<0.01
67	在庫	48	40	8	83.3%	<0.01
68	うえ	24	20	4	83.3%	<0.01
69	どうぞ	24	20	4	83.3%	<0.01
70	おら	12	10	2	83.3%	<0.01
71	残り	12	10	2	83.3%	<0.01
72	振	6	5	1	83.3%	<0.01
73	フリ	6	5	1	83.3%	<0.01
74	R	6	5	1	83.3%	<0.01
75	ペイ	6	5	1	83.3%	<0.01
76	○	6	5	1	83.3%	<0.01
77	添え	6	5	1	83.3%	<0.01
78	ジェネ	6	5	1	83.3%	<0.01
79	7000	6	5	1	83.3%	<0.01
80	ギレ	6	5	1	83.3%	<0.01
81	直	6	5	1	83.3%	<0.01
82	求む	6	5	1	83.3%	<0.01
83	〈アカウント名〉	6	5	1	83.3%	<0.01
84	2seat	6	5	1	83.3%	<0.01
85	引用	6	5	1	83.3%	<0.01
86	アピール	6	5	1	83.3%	<0.01
87	UP	6	5	1	83.3%	<0.01
88	開い	6	5	1	83.3%	<0.01

89	#ミヤカワヒビキ	6	5	1	83.3%	<0.01
90	1mg	47	39	8	83.0%	<0.01
91	くださる	29	24	5	82.8%	<0.01
92	エチ	23	19	4	82.6%	<0.01
93	#レンドルミン	11	9	2	81.8%	<0.01
94	まとめ	11	9	2	81.8%	<0.01
95	〈アカウント名〉	11	9	2	81.8%	<0.01
96	済み	32	26	6	81.3%	<0.01
97	本	31	25	6	80.6%	<0.01
98	ロラゼパム	20	16	4	80.0%	<0.01
99	スクショ	15	12	3	80.0%	<0.01
100	いたし	10	8	2	80.0%	<0.01
101	余っ	157	124	33	79.0%	<0.01
102	到着	19	15	4	78.9%	<0.01
103	ジン	14	11	3	78.6%	<0.01
104	どなた	14	11	3	78.6%	<0.01
105	現状	14	11	3	78.6%	<0.01
106	ジェネリック	41	32	9	78.0%	<0.01
107	値段	41	32	9	78.0%	<0.01
108	150	9	7	2	77.8%	<0.01
109	配送	9	7	2	77.8%	<0.01
110	フリマ	71	55	16	77.5%	<0.01
111	ユーロ	13	10	3	76.9%	<0.01
112	計	13	10	3	76.9%	<0.01
113	or	17	13	4	76.5%	<0.01
114	大阪	17	13	4	76.5%	<0.01
115	確認	58	44	14	75.9%	<0.01
116	ルネスタ	61	46	15	75.4%	<0.01
117	2mg	48	36	12	75.0%	<0.01
118	詳細	16	12	4	75.0%	<0.01
119	入荷	16	12	4	75.0%	<0.01
120	どうしても	16	12	4	75.0%	<0.01
121	認知	12	9	3	75.0%	<0.01
122	エスタゾラム	8	6	2	75.0%	<0.01

123	抽選	8	6	2	75.0%	<0.01
124	800	8	6	2	75.0%	<0.01
125	キノ	8	6	2	75.0%	<0.01
126	フルニト	8	6	2	75.0%	<0.01
127	前日	8	6	2	75.0%	<0.01
128	トリアゾラム	8	6	2	75.0%	<0.01
129	#おくすり	8	6	2	75.0%	<0.01
130	先発	8	6	2	75.0%	<0.01
131	なくなり	8	6	2	75.0%	<0.01
132	テレ	99	74	25	74.7%	<0.01
133	ラボナ	35	26	9	74.3%	<0.01
134	5mg	31	23	8	74.2%	<0.01
135	フルニトラゼパム	58	43	15	74.1%	<0.01
136	フリマアプリ	27	20	7	74.1%	<0.01
137	銀	23	17	6	73.9%	<0.01
138	余り	19	14	5	73.7%	<0.01
139	助かり	19	14	5	73.7%	<0.01
140	お待ち	30	22	8	73.3%	<0.01
141	50	15	11	4	73.3%	<0.01
142	2st	15	11	4	73.3%	<0.01
143	のみ	44	32	12	72.7%	<0.01
144	75	11	8	3	72.7%	<0.01
145	コント	11	8	3	72.7%	<0.01
146	取る	11	8	3	72.7%	<0.01
147	#デエビゴ	40	29	11	72.5%	<0.01
148	PayPay	47	34	13	72.3%	<0.01
149	1000	18	13	5	72.2%	<0.01
150	にて	42	30	12	71.4%	<0.01
151	#ラボナ	35	25	10	71.4%	<0.01
152	売っ	35	25	10	71.4%	<0.01
153	かた	14	10	4	71.4%	<0.01
154	#ジプレキサ	14	10	4	71.4%	<0.01
155	おまけ	7	5	2	71.4%	<0.01
156	100mg	7	5	2	71.4%	<0.01

157	売れる	7	5	2	71.4%	<0.01
158	#thco	7	5	2	71.4%	<0.01
159	信頼	7	5	2	71.4%	<0.01
160	開示	7	5	2	71.4%	<0.01
161	m	17	12	5	70.6%	<0.01
162	エチゾラム	50	35	15	70.0%	<0.01
163	ベンザリン	20	14	6	70.0%	<0.01
164	安価	10	7	3	70.0%	<0.01
165	レボトミン	10	7	3	70.0%	<0.01
166	75mg	10	7	3	70.0%	<0.01
167	18	10	7	3	70.0%	<0.01
168	残	10	7	3	70.0%	<0.01
169	定期	29	20	9	69.0%	<0.01
170	くれる	151	103	48	68.2%	<0.01
171	#ヒルナミン	22	15	7	68.2%	<0.01
172	空	22	15	7	68.2%	<0.01
173	締切	22	15	7	68.2%	<0.01
174	対応	47	32	15	68.1%	<0.01
175	ほしい	53	36	17	67.9%	<0.01
176	ツイ	59	40	19	67.8%	<0.01
177	お願い	197	133	64	67.5%	<0.01
178	#ゾルピデム	45	30	15	66.7%	<0.01
179	プレゼント	42	28	14	66.7%	<0.01
180	その他	24	16	8	66.7%	<0.01
181	物	12	8	4	66.7%	<0.01
182	1s	12	8	4	66.7%	<0.01
183	実績	12	8	4	66.7%	<0.01
184	詳しく	12	8	4	66.7%	<0.01
185	サインバルタ	9	6	3	66.7%	<0.01
186	#抗うつ剤	9	6	3	66.7%	<0.01
187	パ	23	15	8	65.2%	<0.01
188	ごさい	77	50	27	64.9%	<0.01
189	問い合わせ	17	11	6	64.7%	<0.01
190	報告	45	29	16	64.4%	<0.01

191	paypay	25	16	9	64.0%	<0.01
192	買い	47	30	17	63.8%	<0.01
193	発送	58	37	21	63.8%	<0.01
194	10	80	51	29	63.8%	<0.01
195	#クエチアピン	11	7	4	63.6%	<0.01
196	#プレガバリン	11	7	4	63.6%	<0.01
197	下さい	182	115	67	63.2%	<0.01
198	いらっしゃい	19	12	7	63.2%	<0.01
199	シート	100	63	37	63.0%	<0.01
200	♂	27	17	10	63.0%	<0.01
201	全て	27	17	10	63.0%	<0.01
202	活動	35	22	13	62.9%	<0.01
203	おり	35	22	13	62.9%	<0.01
204	以外	43	27	16	62.8%	<0.01
205	余る	16	10	6	62.5%	<0.01
206	大丈夫	61	38	23	62.3%	<0.01
207	居	42	26	16	61.9%	<0.01
208	フル	26	16	10	61.5%	<0.01
209	#DM ください	13	8	5	61.5%	<0.01
210	ハルシオン	13	8	5	61.5%	<0.01
211	20mg	13	8	5	61.5%	<0.01
212	デパ	36	22	14	61.1%	<0.01
213	#フルニトラゼパム	58	35	23	60.3%	<0.01
214	安い	30	18	12	60.0%	<0.01
215	ミン	20	12	8	60.0%	<0.01
216	程度	15	9	6	60.0%	<0.01
217	#お薬譲ります	254	152	102	59.8%	<0.01
218	欲しい	184	110	74	59.8%	<0.01
219	いら	32	19	13	59.4%	<0.01
220	箱	22	13	9	59.1%	<0.01
221	アプリ	12	7	5	58.3%	<0.01
222	いただき	12	7	5	58.3%	<0.01
223	宜しく	12	7	5	58.3%	<0.01
224	早め	12	7	5	58.3%	<0.01

225	サイレ	12	7	5	58.3%	<0.01
226	デ	248	144	104	58.1%	<0.01
227	#眠剤	69	40	29	58.0%	<0.01
228	プレガバリン	19	11	8	57.9%	<0.01
229	ミリ	47	27	20	57.4%	<0.01
230	メルカリ	47	27	20	57.4%	<0.01
231	ゾルピデム	35	20	15	57.1%	<0.01
232	ソラナックス	28	16	12	57.1%	<0.01
233	デパケン	14	8	6	57.1%	<0.01
234	300	14	8	6	57.1%	<0.01
235	可能	62	35	27	56.5%	<0.01
236	パス	224	126	98	56.3%	<0.01
237	25mg	16	9	7	56.3%	<0.01
238	送れ	16	9	7	56.3%	<0.01
239	#安定剤	16	9	7	56.3%	<0.01
240	ワイパックス	27	15	12	55.6%	<0.01
241	デエビゴ	47	26	21	55.3%	<0.01
242	中身	20	11	9	55.0%	<0.01
243	致し	51	28	23	54.9%	<0.01
244	レース	236	129	107	54.7%	<0.01
245	コンサータ	86	47	39	54.7%	<0.01
246	サイ	291	158	133	54.3%	<0.01
247	企画	65	35	30	53.8%	<0.01
248	梱包	26	14	12	53.8%	<0.01
249	出来る	28	15	13	53.6%	<0.01
250	大量	60	32	28	53.3%	<0.01
251	クエチアピン	32	17	15	53.1%	<0.01
252	10mg	36	19	17	52.8%	<0.01
253	05	21	11	10	52.4%	<0.01
254	固定	23	12	11	52.2%	<0.01
255	待つ	25	13	12	52.0%	<0.01
256	フォロー	66	33	33	50.0%	<0.01
257	ありがとう	44	22	22	50.0%	<0.01
258	#リボトリール	32	16	16	50.0%	<0.01

259	要注意	32	16	16	50.0%	<0.01
260	レンドルミン	22	11	11	50.0%	<0.01
261	#ロヒプノール	20	10	10	50.0%	<0.01
262	ください	514	254	260	49.4%	<0.01
263	リリカ	76	37	39	48.7%	<0.01
264	書い	29	14	15	48.3%	<0.01
265	困っ	25	12	13	48.0%	<0.01
266	取引	111	53	58	47.7%	<0.01
267	レキソタン	49	23	26	46.9%	<0.01
268	#リリカ	47	22	25	46.8%	<0.01
269	ベルソムラ	28	13	15	46.4%	<0.01
270	レタパ	39	18	21	46.2%	<0.01
271	かなり	26	12	14	46.2%	<0.01
272	気軽	42	19	23	45.2%	<0.01
273	よろしく	89	40	49	44.9%	<0.01
274	♀	65	29	36	44.6%	<0.01
275	1st	46	20	26	43.5%	<0.01
276	買っ	37	16	21	43.2%	<0.01
277	20	44	19	25	43.2%	<0.01
278	眠	124	53	71	42.7%	<0.01
279	できる	82	35	47	42.7%	<0.01
280	スリー	203	86	117	42.4%	<0.01
281	募集	45	19	26	42.2%	<0.01
282	リボトリール	39	16	23	41.0%	<0.01
283	剤	151	61	90	40.4%	<0.01
284	#サイレース	274	106	168	38.7%	<0.01
285	ご	218	84	134	38.5%	<0.01
286	2	137	52	85	38.0%	<0.01
287	#おくすりもぐもぐ	162	60	102	37.0%	<0.01
288	相談	102	37	65	36.3%	<0.01
289	連絡	292	100	192	34.2%	<0.01
290	1	214	73	141	34.1%	<0.01
291	#デパス	497	144	353	29.0%	<0.01
292	#マイスリー	334	95	239	28.4%	<0.01



293	#お薬もぐもぐ	6682	1641	5041	24.6%	<0.01
294	てる	737	142	595	19.3%	<0.01
295	たら	747	127	620	17.0%	<0.01
296	けど	456	74	382	16.2%	<0.01
297	れ	478	75	403	15.7%	<0.01
298	もぐもぐ	370	57	313	15.4%	<0.01
299	き	210	32	178	15.2%	<0.01
300	垢	278	42	236	15.1%	<0.01
301	聞き	7	1	6	14.3%	<0.01
302	ば	316	45	271	14.2%	<0.01
303	つけ	142	18	124	12.7%	<0.01
304	明日	136	17	119	12.5%	<0.01
305	今日	215	26	189	12.1%	<0.01
306	情報	100	12	88	12.0%	<0.01
307	だろ	78	9	69	11.5%	<0.01
308	師	128	14	114	10.9%	<0.01
309	出	76	8	68	10.5%	<0.01
310	なっ	232	23	209	9.9%	<0.01
311	かも	92	9	83	9.8%	<0.01
312	なあ	175	17	158	9.7%	<0.01
313	すぎる	83	8	75	9.6%	<0.01
314	または	64	6	58	9.4%	<0.01
315	#レキソタン	161	15	146	9.3%	<0.01
316	プロフ	451	39	412	8.6%	<0.01
317	友達	70	6	64	8.6%	<0.01
318	ブロン	72	6	66	8.3%	<0.01
319	もらっ	75	6	69	8.0%	<0.01
320	ちょっと	50	4	46	8.0%	<0.01
321	アカウント	76	6	70	7.9%	<0.01
322	#精神疾患	51	4	47	7.8%	<0.01
323	増え	52	4	48	7.7%	<0.01
324	しよ	67	5	62	7.5%	<0.01
325	飲み	124	9	115	7.3%	<0.01
326	量	42	3	39	7.1%	<0.01

327	処方箋	28	2	26	7.1%	<0.01
328	飲む	72	5	67	6.9%	<0.01
329	メッセージ	234	16	218	6.8%	<0.01
330	病院	105	7	98	6.7%	<0.01
331	皆様	45	3	42	6.7%	<0.01
332	ふわふわ	46	3	43	6.5%	<0.01
333	すぐ	46	3	43	6.5%	<0.01
334	追加	77	5	72	6.5%	<0.01
335	RT	2037	132	1905	6.5%	<0.01
336	使わ	49	3	46	6.1%	<0.01
337	限り	67	4	63	6.0%	<0.01
338	頭	34	2	32	5.9%	<0.01
339	気持ち	34	2	32	5.9%	<0.01
340	LINE	52	3	49	5.8%	<0.01
341	#詐欺師	52	3	49	5.8%	<0.01
342	ながら	70	4	66	5.7%	<0.01
343	科	35	2	33	5.7%	<0.01
344	#OD	263	15	248	5.7%	<0.01
345	Twitter	107	6	101	5.6%	<0.01
346	海外	54	3	51	5.6%	<0.01
347	紹介	54	3	51	5.6%	<0.01
348	良い	92	5	87	5.4%	<0.01
349	東京	37	2	35	5.4%	<0.01
350	注意	351	18	333	5.1%	<0.01
351	向け	39	2	37	5.1%	<0.01
352	こんな	80	4	76	5.0%	<0.01
353	寝る	40	2	38	5.0%	<0.01
354	生き	40	2	38	5.0%	<0.01
355	思う	40	2	38	5.0%	<0.01
356	思い	370	18	352	4.9%	<0.01
357	思っ	62	3	59	4.8%	<0.01
358	でき	444	21	423	4.7%	<0.01
359	#病み垢	127	6	121	4.7%	<0.01
360	飲ん	234	11	223	4.7%	<0.01

361	問題	48	2	46	4.2%	<0.01
362	#ブロン	235	9	226	3.8%	<0.01
363	24	210	8	202	3.8%	<0.01
364	被害	106	4	102	3.8%	<0.01
365	個人	213	8	205	3.8%	<0.01
366	本人	27	1	26	3.7%	<0.01
367	#メジコン	164	6	158	3.7%	<0.01
368	販売	83	3	80	3.6%	<0.01
369	配信	28	1	27	3.6%	<0.01
370	メンヘラ	28	1	27	3.6%	<0.01
371	譲渡	28	1	27	3.6%	<0.01
372	人気	85	3	82	3.5%	<0.01
373	一緒	29	1	28	3.4%	<0.01
374	クリ	30	1	29	3.3%	<0.01
375	OD	363	12	351	3.3%	<0.01
376	ましよ	122	4	118	3.3%	<0.01
377	譲る	31	1	30	3.2%	<0.01
378	メン	32	1	31	3.1%	<0.01
379	日本	65	2	63	3.1%	<0.01
380	行っ	99	3	96	3.0%	<0.01
381	依存	66	2	64	3.0%	<0.01
382	#拡散希望	33	1	32	3.0%	<0.01
383	ダメ	33	1	32	3.0%	<0.01
384	今週	34	1	33	2.9%	<0.01
385	処方	419	12	407	2.9%	<0.01
386	リスク	36	1	35	2.8%	<0.01
387	防止	38	1	37	2.6%	<0.01
388	中途	38	1	37	2.6%	<0.01
389	仲良く	39	1	38	2.6%	<0.01
390	悪い	39	1	38	2.6%	<0.01
391	夜	353	9	344	2.5%	<0.01
392	#病み垢さんと繋がりたい	503	12	491	2.4%	<0.01
393	起き	42	1	41	2.4%	<0.01
394	として	42	1	41	2.4%	<0.01

395	手続き	42	1	41	2.4%	<0.01
396	違法	86	2	84	2.3%	<0.01
397	について	132	3	129	2.3%	<0.01
398	朝	132	3	129	2.3%	<0.01
399	証	44	1	43	2.3%	<0.01
400	払っ	44	1	43	2.3%	<0.01
401	覚醒	45	1	44	2.2%	<0.01
402	売買	45	1	44	2.2%	<0.01
403	うわ	45	1	44	2.2%	<0.01
404	#精神科	46	1	45	2.2%	<0.01
405	全く	331	7	324	2.1%	<0.01
406	仕事	99	2	97	2.0%	<0.01
407	辛い	50	1	49	2.0%	<0.01
408	保険	51	1	50	2.0%	<0.01
409	受け	103	2	101	1.9%	<0.01
410	られる	103	2	101	1.9%	<0.01
411	喚起	208	4	204	1.9%	<0.01
412	#病み垢女子	55	1	54	1.8%	<0.01
413	ぜひ	234	4	230	1.7%	<0.01
414	#薬物	61	1	60	1.6%	<0.01
415	返事	187	3	184	1.6%	<0.01
416	OK	259	4	255	1.5%	<0.01
417	不眠症	131	2	129	1.5%	<0.01
418	管理	333	5	328	1.5%	<0.01
419	#覚醒剤	67	1	66	1.5%	<0.01
420	懲役	71	1	70	1.4%	<0.01
421	病み	234	3	231	1.3%	<0.01
422	薬局	78	1	77	1.3%	<0.01
423	特に	314	4	310	1.3%	<0.01
424	お知らせ	161	2	159	1.2%	<0.01
425	乱用	84	1	83	1.2%	<0.01
426	輸入	172	2	170	1.2%	<0.01
427	あやしい	101	1	100	1.0%	<0.01
428	ヤクヅツ	101	1	100	1.0%	<0.01

429	ネット	109	1	108	0.9%	<0.01
430	入り	773	7	766	0.9%	<0.01
431	#個人輸入	114	1	113	0.9%	<0.01
432	ホームページ	115	1	114	0.9%	<0.01
433	#大麻	121	1	120	0.8%	<0.01
434	#メンヘラ	858	7	851	0.8%	<0.01
435	精神	517	4	513	0.8%	<0.01
436	医療	397	3	394	0.8%	<0.01
437	入れ	811	6	805	0.7%	<0.01
438	現実	313	2	311	0.6%	<0.01
439	食べ	318	2	316	0.6%	<0.01
440	#DXM	165	1	164	0.6%	<0.01
441	薬物	165	1	164	0.6%	<0.01
442	リンク	396	2	394	0.5%	<0.01
443	タグ	420	2	418	0.5%	<0.01
444	#メンヘラさん とつながり たい	801	3	798	0.4%	<0.01
445	放置	325	1	324	0.3%	<0.01
446	転売	329	1	328	0.3%	<0.01
447	地獄	349	1	348	0.3%	<0.01
448	#睡眠障害	359	1	358	0.3%	<0.01
449	クスリ	370	1	369	0.3%	<0.01
450	#オーバードーズ	424	1	423	0.2%	<0.01
451	向	445	1	444	0.2%	<0.01
452	#メンヘラさん と繋がりたい	796	1	795	0.1%	<0.01

Table 2. 機械学習で作成したモデルの判別能力結果

モデル	データ区分	テスト		学習		検証	
		件数	割合	件数	割合	件数	割合
7:3	正解	1,867	82.4%	4,206	80.4%		
	誤り	401	17.7%	1,025	19.6%		
6:2:2	正解	1,301	83.8%	3,564	80.1%	1,208	80.8%
	誤り	251	16.2%	888	20.0%	287	19.2%

Table 3. 機械学習で作成したモデルの検証-評価メトリック

モデル	テスト		学習		検証	
	AUC	Gini	AUC	Gini	AUC	Gini
7:3	0.838	0.676	0.825	0.650		
6:2:2	0.842	0.684	0.824	0.648	0.834	0.668

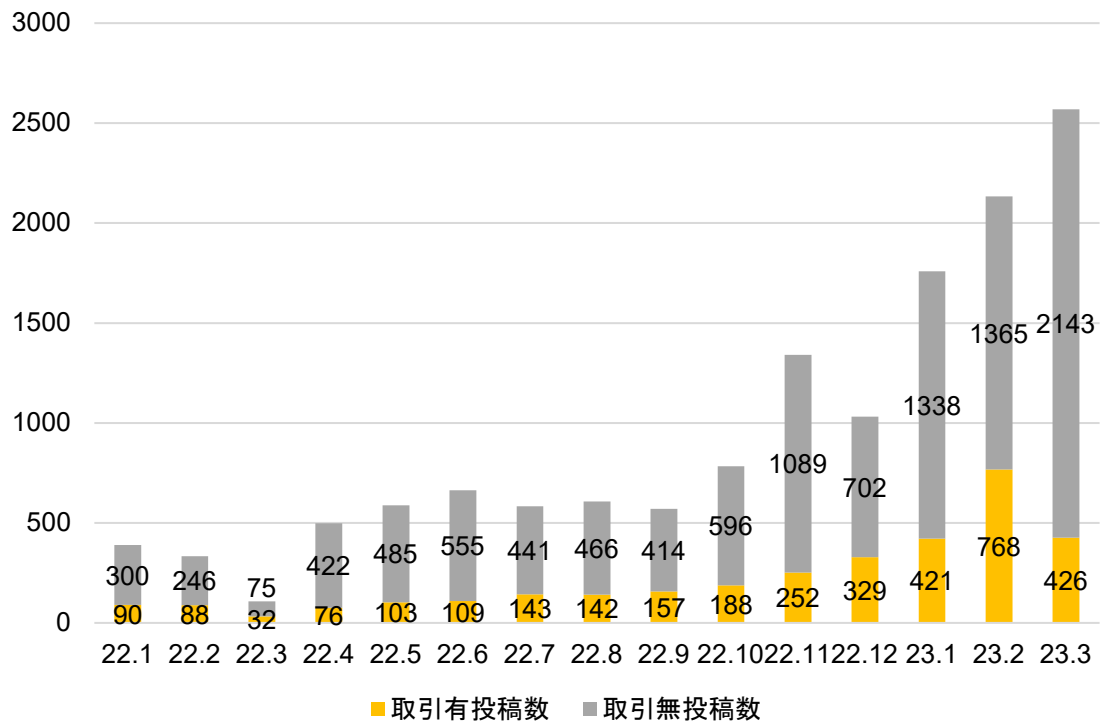


Figure 1. 月別投稿数

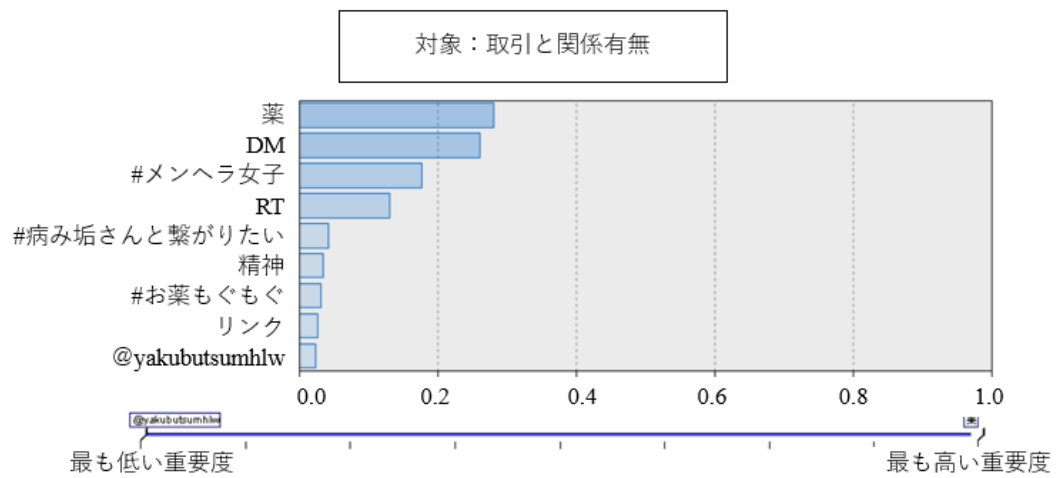


Figure 2. 予測変数の重要度 (学習：テスト=7:3)

2022 年データセットを用いた CHAID による医薬品の個人間取引疑い有無の予測モデルにおける予測変数重要度

\* : 0.0-1.0 は重要度、数値は大きいほど、重要度が高くなる。

\*



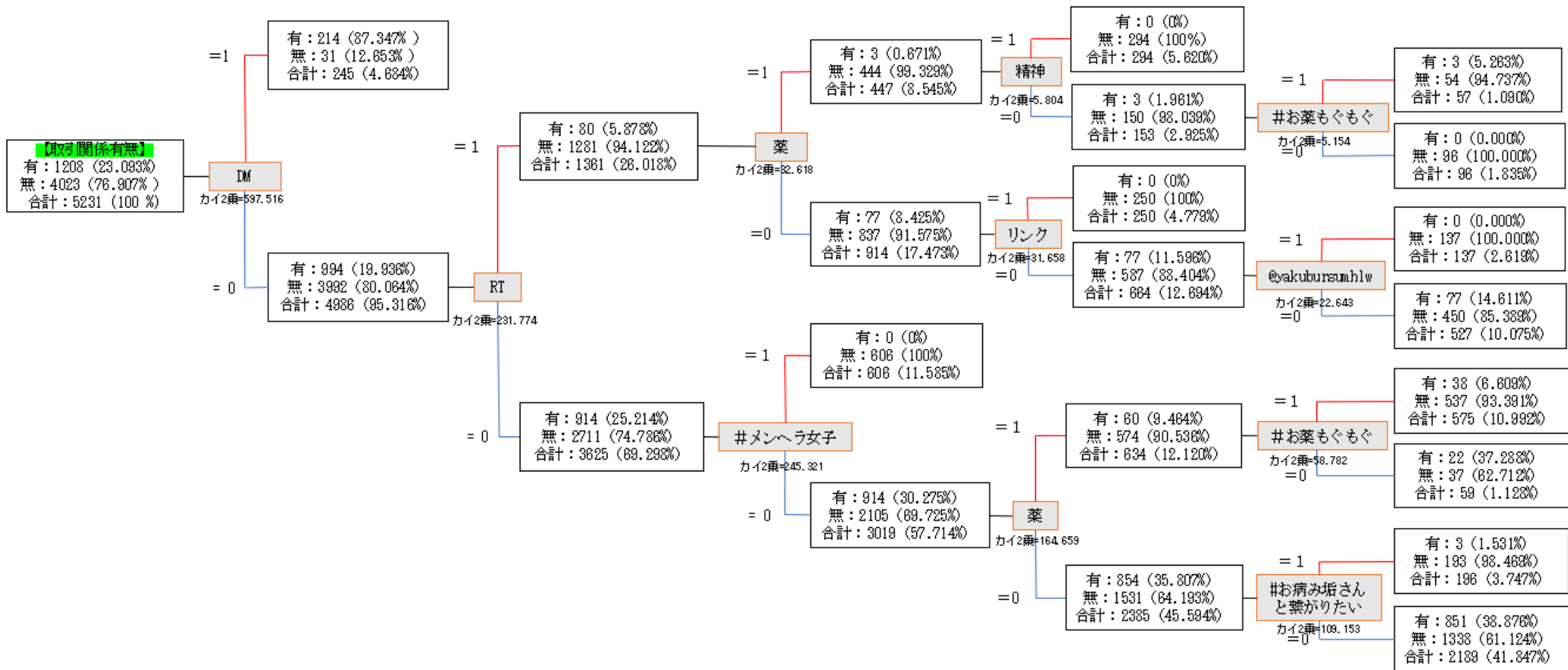


Figure 3.生成した決定木モデル 2022 (学習 : テスト = 7:3)

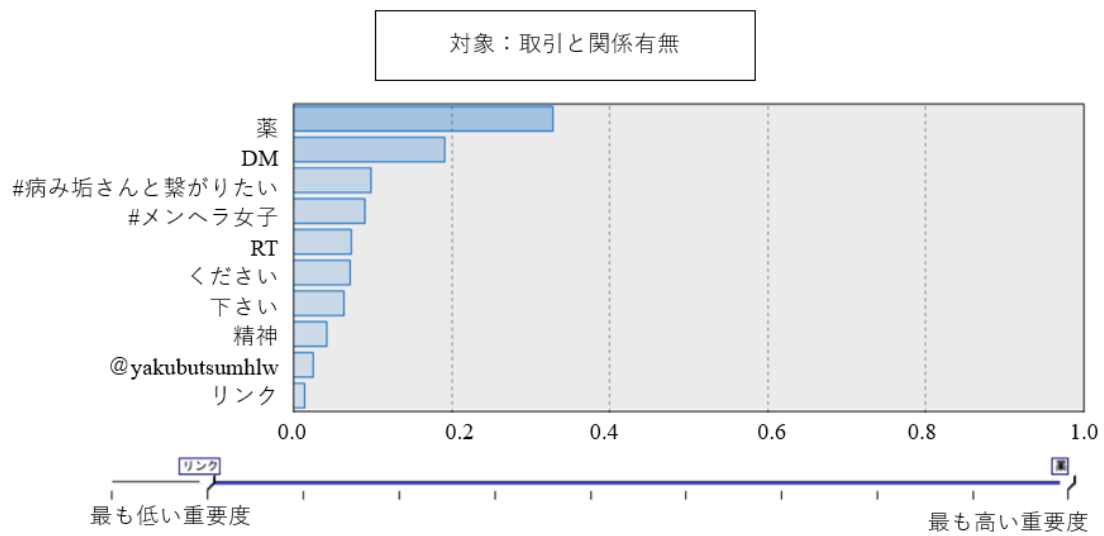


Figure 4. 予測変数の重要度 (学習：テスト：検証=6：2：2)

2022 年データセットを用いた CHAID による医薬品の個人間取引疑い有無の予測モデルにおける予測変数重要度

\*：0.0-1.0 は重要度、数値は大きいほど、重要度が高くなる。

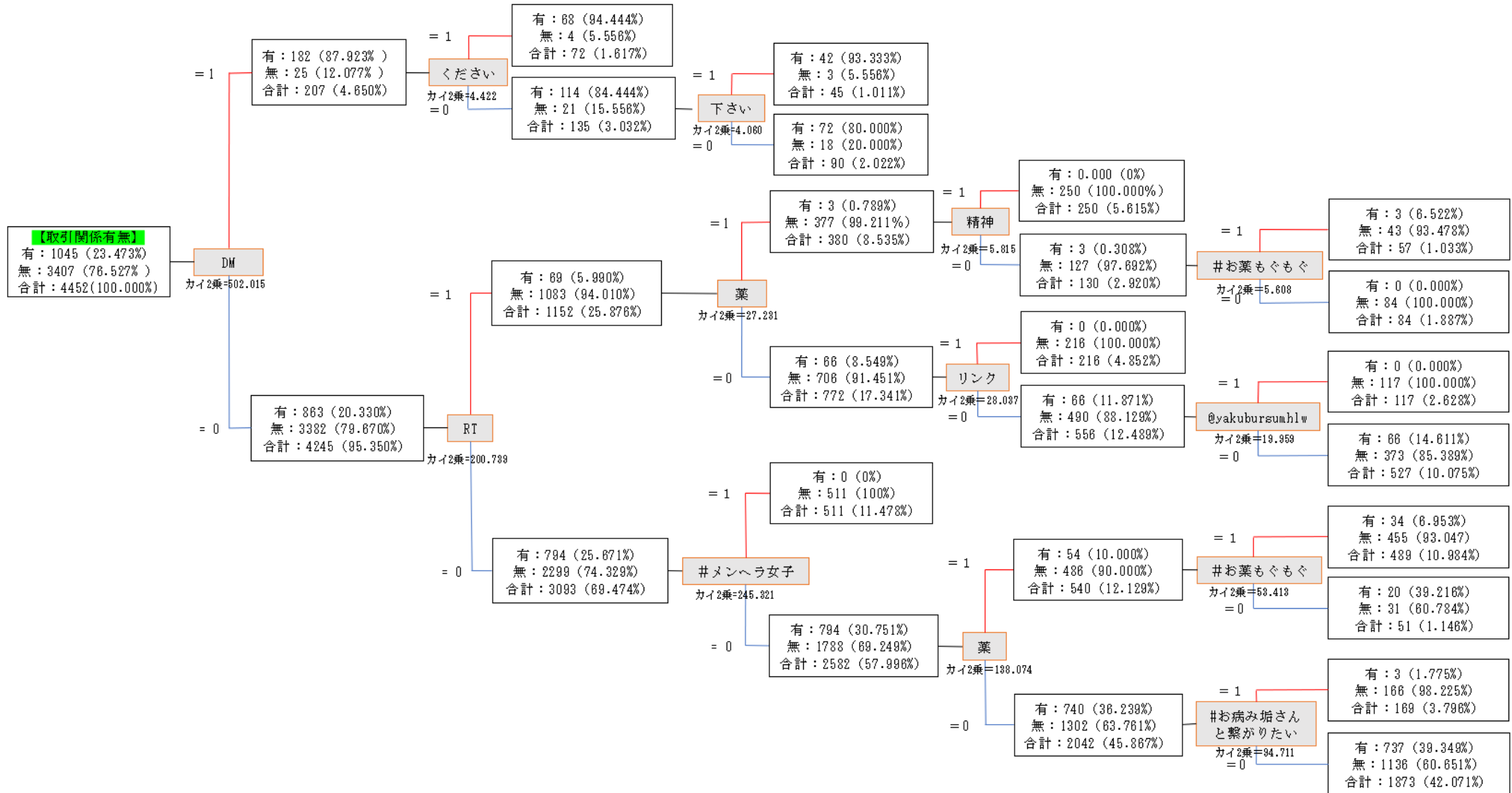


Figure 5. 生成した決定木モデル 2022 (学習 : テスト : 検証 = 6 : 2 : 2)

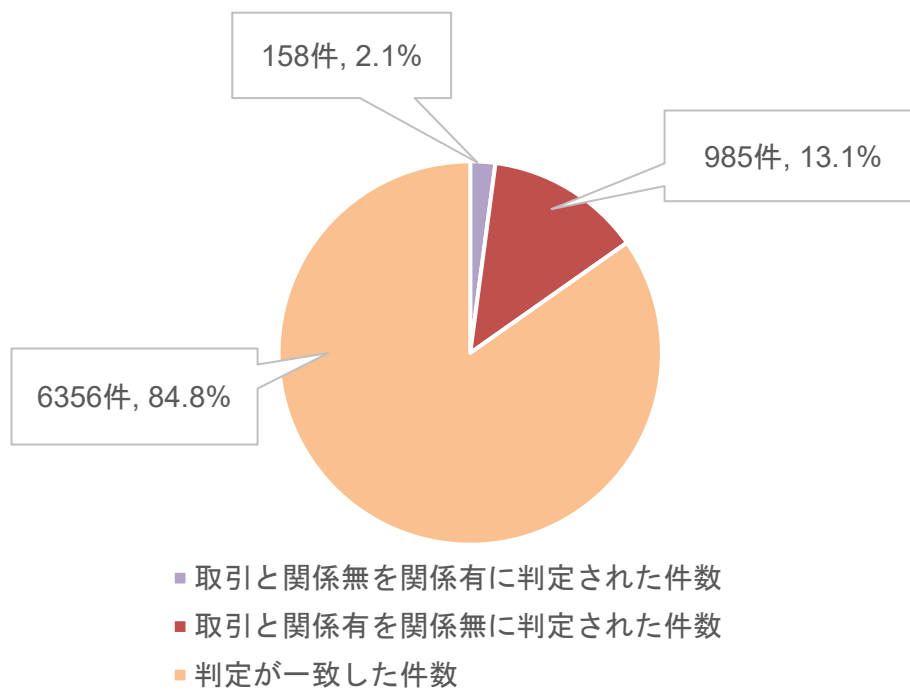


Figure 6. 2023 年のデータで予測モデル（学習：テスト＝7:3）を検証した結果