新機能紹介

NTT DATA
Trusted Global Innovator

TMS_beta（テキスト処理ベータ版）

のご紹介

株式会社NTTデータ数理システム

TMS_beta（テキスト処理ベータ版）とは

テキスト処理（ベータ版）は当社製品の分析プラットフォームMSIP上でテキストデータを分析するための機能群です。

こんな機能をご提供いたします！

1. 自由記述の文章を機械的に扱えるようにする「分かち書き」や各種分析機能に適用するための「フィルタリング」などの**テキスト処理機能のアイコン群**
2. MSIP上でテキスト処理を行う作業をサポートする**テクニカルサンプルプロジェクト**

こんなことが実現できます！

1. アンケートデータからよくあるご要望を抽出する
2. 不具合情報やニュース記事などの分類・予測を行う（Alkano連携）

利用イメージ

テキストデータを「分かち書き」「フィルタリング」したうえで、集計等の基礎分析や統計解析・機械学習の分析手法を組み合わせて利用することで、様々な目的に沿った分析をしていただけます。

特長１：テキストマイニングに使いやすい「分かち書き」

TMS_betaの「分かち書き」機能は、ただ単に単語に区切ってその品詞や原形を出力するだけではありません。辞書登録なしでちょうどいい単位で単語を抽出する「**自動連結**」や、「**係り受け関係**」「**態度表現**」を自動で取得することで、一歩進んだテキストマイニングを実現します。

- テキストマイニングをするうえで、ちょうどいい単位で自動的に単語を切り出せます。
  （たとえば上の例では、「東京スカイツリー」も、辞書登録なしで扱えます。）
- **係り受け関係**や**態度表現**の情報を取得することで、**意味を考慮した分析**も可能です。

特長２：わかりやすいGUIでルールベースのラベリングを支援する「カテゴリ生成」

TMS_beta では、テキストデータに対してルールベースでラベル（カテゴリ）を付与する「**カテゴリ生成**」アイコンを提供します。わかりやすいGUIで、特定の観点や話題でテキストをまとめあげるためのルールを定義し、観点・話題のカテゴリを作成することができます。テキストの概要把握、各種機械学習の入力データ（説明変数の利用）、分類・予測分析における教師データの作成などに有効にご活用いただけます。

入力（テキストデータ）※

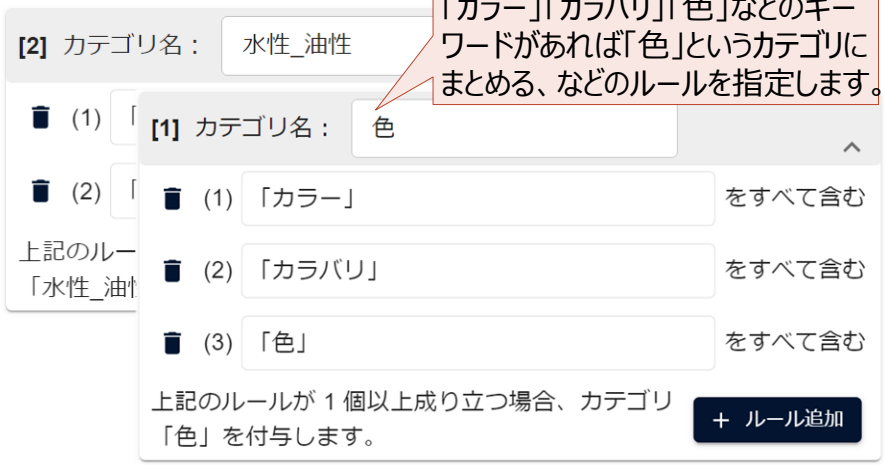
- いろんな種類のカラーを使えること。仕事用の油性ペンとプライベート用の4色ペンをいつも持ち歩いています
- ノートをきれいにまとめる用にパステル系のカラーがたくさん入ったものを選んでます。あと細字の方がかわいくていい。
- スムーズに書けるものがいいですね。太字のもののほうが見やすいです。
- 軽くてサラサラと書ける水性ペンをずっと使っています。

⋮

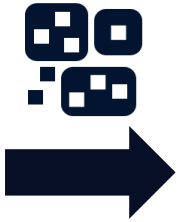
※実際には、「分かち書き」アイコンで分かち書きをした後のデータを入力とします。

カテゴリルール

ルールを決める
「カラー」「カラバリ」「色」などのキーワードがあれば「色」というカテゴリにまとめる、などのルールを指定します。



カテゴリ生成



出力（カテゴリチェック表）

テキスト	色	水性_油性
いろんな種類の**カラー**を使えること。仕事用の**油性**ペンとプライベート用の4色ペンをいつも持ち歩いています	true	true
ノートをきれいにまとめる用にパステル系の**カラー**がたくさん入ったものを選んでます。あと細字の方がかわいくていい。	true	false
スムーズに書けるものがいいですね。太字のもののほうが見やすいです。	false	false
軽くてサラサラと書ける**水性**ペンをずっと使っています。	false	true

指定したルールに沿って単語やテキストをまとめる

話題・観点のキーワードがテキスト中に現れていれば true, なければ false という値の列が追加されます。話題・観点の有無のデータ（カテゴリデータ）を作成します。

TMS_beta（テキスト処理ベータ版）の機能一覧

分かち書き

テキスト（文章）を単語や文節単位に分割し、品詞や係り受け（単語の意味的なつながり）の情報を解析します。
人が見て解釈できる単位での文節を判定する「自動連結」機能や、否定や要望など記述者の主観等を表現する「態度表現」を付与する機能を有しています。

分かち書き結果のフィルタリング

品詞や単語の出現頻度、文字列や文字数によるフィルタリングを行い、分析に有用な単語を抽出します。
【フィルタリング例】
品詞フィルタ：名詞や動詞など、単体で意味を持つ単語を抽出します。
頻度フィルタ：頻度2以上かつ頻度上位5件を除外など、程よくまんべんなく使われている単語を抽出します。
文字列フィルタ：製品名の一部など、「この文字列を含む単語」とターゲットを絞ることができます。
文字数フィルタ：3文字以上と設定し、複合語などより「専門用語」らしい単語を抽出します。

重み算出

属性（テキストに紐づく付加情報）と組み合わせ、属性ごとの単語の重み（重要度）を算出します。
属性ごとに重みが大きな単語を見ることで属性の傾向を把握することができます（TMS特徴分析結果相当）。
さらに重みの値を各種機械学習の入力（説明変数）として分類・予測に利用いただくことも可能です。

テキストカテゴリ化

テキストデータ内に出現する単語や係り受け表現のキーワードをもとにテキストデータのカテゴリ化（グループ分け）を行います。
※詳細は p.4 をご覧ください。

テクニカルサンプルプロジェクト

Alkano のワークフローは、自分で分析を組み立てていかなければならないので大変そう、というご心配のお声をよく聞きます。そこでテキストデータの分析を行う**分析ワークフローのサンプル**とそのプロジェクトの**解説資料（pdf）**をご用意いたしました。(https://www.msi.co.jp/solution/msip/samples/index.html)

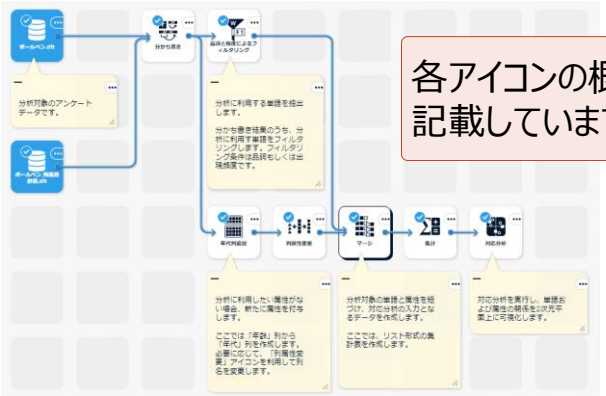
こんな方にお勧めです！

- Alkano (MSIP) を使ったテキスト分析をはじめて行う方
- Alkano(MSIP)で数値・カテゴリデータと合わせて、テキストデータの分析を始めたい方
- これまで Text Mining Studio を利用されていて、分析フローを組み立てるのが初めての方

テクニカルサンプルプロジェクトの特徴

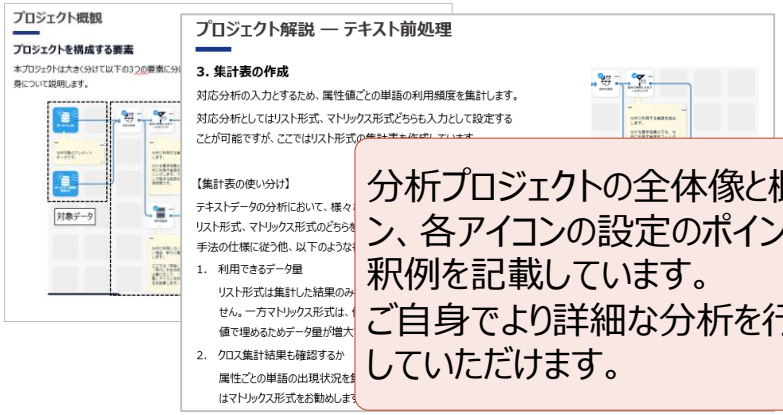
- サンプルデータをお手元のデータに差し替えてご利用いただくことで、データ分析を始めやすい
- 説明資料には分析・設定のポイントを明記しているので、お手元での試行錯誤のポイントが分かりやすい

▼ 分析フローのプロジェクト



各アイコンの概要をメモに記載しています。

▼ 分析フローの解説資料



分析プロジェクトの全体像と概要、活用シーン、各アイコンの設定のポイントや結果の解釈例を記載しています。
ご自身でより詳細な分析を行う際の参考にしていただけます。

TMS_beta（テキスト処理ベータ版）のご利用につきまして

- TMS_betaは、当社製品 MSIP 上で Text Mining Studio のテキスト処理機能群をご利用いただくための第1ステップとして実現した機能を、ベータ版として無料でご提供するものです。
- Alkano または BayoLinkS の保守をご契約中のお客様（一部のライセンスを除く※）は、TMS_betaを **2024 年 3 月 31 日まで** 無料でご利用いただけます。無料期間終了後も継続してご利用をご希望の場合には、有料でのご提供を予定しております。（有料でのご利用にはお手続きが必要となります。何もお手続きいただかない場合は無料期間終了後は、自動でご利用いただけなくなります。）
  - Alkano または BayoLinkS を 2023 年 3 月 27 日以降にご契約いただいたお客様につきましては、MSIPがインストールされている環境であれば、別途インストールの必要はございませんが、別紙「ユーザー登録フォーム」の「TMS_beta」の欄にチェックを入れた方のみご利用いただけます。
  - すでに Alkano または BayoLinkS をご契約中のお客様で TMS_betaのご利用をご希望の方は別紙「TMS_betaお申込書」にてお申込みください。TMS_betaのご利用にはAlkanoまたはBayoLinkSを最新版にバージョンアップいただく必要がございます。また、お申し込み受付終了後に、ライセンスの更新作業も必要になります。（詳細は当社よりご案内差し上げます）

※Alkano Standard Simple/Alkano SA Simple/BayoLinkS Basic (Simple)/ BayoLinkS Standard (Simple)をご利用の方はTMS_betaをご利用いただけません。ご利用をご希望の場合は、当社までお問い合わせください。

お問い合わせ先

株式会社NTTデータ数理システム　営業部　東京都新宿区信濃町35番地 信濃町煉瓦館1階
Phone：03-3358-6681
E-mail：vmstudio-support@ml.msi.co.jp
受付時間は、当社Webサイト（https://www.msi.co.jp/contact/index.html）をご確認ください。

NTT DATA

Trusted Global Innovator