

3

AI-OCR用いた輸出貿易
統計のデータセット化
(1970~1987年、1988~2019年)

AI-OCR用いた輸出貿易統計のデータセット化

1970～2019年の農林水産物及び調味料・飲料等の貿易統計情報について、1970～1987年、1988年～2019年、それぞれ下記方法にてデータセット化を実施

対象年	1970～1987年	1988年～2019年
ソース	https://www.customs.go.jp/toukei/suii/index.htm	https://www.customs.go.jp/toukei/srch/index.htm?M=01&P=0
対象品目	・ 01類～05類: 動物及び動物性生産品(含魚介類) ・ 06類～15類: 植物性生産品、動植物油脂 ・ 16類～24類: 調製食料品、飲料、アルコール、たばこ等	同左
データセット化方法	・ 年ごとにAI-OCRにてデータセット化 ・ 精度は97.7%程度 ※品目名を除く	・ 年ごとに上記URLよりCSVをダウンロード ・ 精度は100%程度

4

AIコンペの開催支援

1. AIコンペ募集方法
2. AIコンペ参加候補者情報
3. AIコンペの企画案
(コンペの課題、審査基準、スケジュール等)
4. AIコンペ募集サイト
5. AI/アナリティクスを活用した輸出分析及び政策立案
に有効な情報リスト

1. AIコンペ募集方法

AIコンペの目的・ゴール及び前提/制約

AI/アナリティクスにより解くべき課題の設定・具体化及び農林水産省のAI取組みをより多くの日人に周知し、AI人材をより多く集めることをゴールとする

目的	✓ 大量かつ多様なデータな分析及び分析に基づく新たな政策手法の導出(EBPMの推進) ✓ AI等の新たな技術の業務活用を検討するうえでの実証実験パートナー候補の具体化
ゴール	✓ 今回企画するAIコンペにて解決する課題(⇒コンペのテーマ設定) ・ 上記目的の達成に向けたAI/アナリティクスの活用可能性の具体化 ⇒AI/アナリティクスにより解くべき課題の設定 ✓ 実証実験パートナーの具体化対象・数(⇒コンペのテーマ設定・期間・募集方法等) ・ 質よりも数:有象無象含めたロングリストの作成
前提/制約	■ 時間軸 : 制約なし。次年度の開催も可 ■ 一般公開 : 公開したい。農林水産省のAI取組みの周知目的もあり ■ 特典 : 農林中金等のスポンサーも検討 ■ 共催 : 可能?

AIコンペの公示・募集方法

今回の目的・ゴールを鑑み、既にAIコンペの実績・会員を有するAIコンペ公募サイト、中でも実績が他を圧倒するSIGNATE社のサイトでの募集を検討する

	各社への個別打診	一般公募
概要	■ 下記リレーション・リスト等で打診 ✓ PwCリレーション (ベンチャー自身、VC) ✓ METI大学発ベンチャーリスト ✓ 東京大学産業協創推進本部支援先 ✓ 他省庁コンペ受賞企業 ✓ その他ベンチャー支援組織 ・DEEPCORE ・Plug & Play ・Wantedly(求人情報)など	■ <u>貴省のHPにて一般公募</u> + コミュニティサイトへの掲示 例) 人工知能研究コミュニティ AI Academy + ベンチャー支援組織への声掛け依頼 + 各社個別打診
メリット	✓ コンペの設定が緩くても可 ✓ 打診過程で一定レベルで優秀さ・機動力の確認が可能	✓ コンペの設定が緩くても可 ✓ 公示(透明性)あり
デメリット	✓ リストの数は絞られる	✓ 認知力弱い⇒応募数少なる可能性あり ✓ 評価・絞込みが困難な可能性あり

■ AIコンペサイトでの公募

SIGNATE

設立: 2017年5月

開催コンペ数: 約50

会員数: 28000人

経産省・国立図書館等の
官公庁実績あり

- ✓ ProbSpace
設立: 2018年4月設立
開催コンペ数: 4件
会員数: 不明
- ✓ Nishika
設立: 2019年5月
開催コンペ数: 1件
会員数: 不明

- ✓ 認知力あり⇒応募数の確保は容易
- ✓ 公示(透明性)あり

- ✓ コンペの設定の詳細化・厳格化が必要
- ✓ 評価・絞込み(一次)が困難

2. AIコンペ参加候補者情報

AIコンペ参加候補者

AIコンペティションの候補者は、日本最大のAI開発人材の登録を有するSIGNATEの会員を対象に告知

2020年3月18日現在の登録会員数

28,142名

社会人：20,486名 73%

学生：7,656名 27%

コンペ参加人数：15,536名 55%

会員属性

理工・情報系卒、あるいは、最終学歴が修士・博士といったAI・データサイエンスに強い人材が会員の過半数を占める

プログラミング言語

Python&Rユーザ

90%



ステータス

社会人

73%

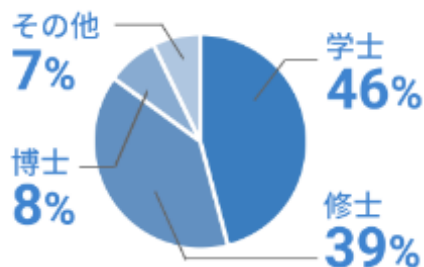


学生

27%



最終学歴



卒業学部・学科

理工・情報系卒



65%

▼会員の主な出身大学

- ・ 東京大学
- ・ 早稲田大学
- ・ 慶應義塾大学
- ・ 東京理科大学
- ・ 筑波大学
- ・ 東京工業大学
- ・ 大阪大学
- ・ 大阪府立大学
- ・ 電気通信大学
- ・ 明治大学
- ・ 京都大学
- ・ 東北大学
- ・ 横浜国立大学
- ・ 同志社大学
- ・ 中央大学
- ・ 青山学院大学 等

国、行政コンペ実績

全てのコンペティションにて、企業や研究機関が持つ既存ベンチマークを大幅に上回る結果を取得

クライアント名	事業名
一般財団法人 日本情報経済社会推進協会(JIPDEC)	「IoTを活用した新産業モデル創出基盤整備事業／革新的な産業モデル創出及び推進に関する検討」のうち、「ビッグデータ分析コンテストの実施に必要な支援」 (第三回・第四回)
産業技術総合研究所(AIST)	AIコンテスト企画・運営業務
沖縄地理情報システム協議会(GIS)	平成30年度 内閣府沖縄型産業中核人材育成事業「地理空間ビッグデータ活用人材育成プログラム」
	令和元年度 内閣府沖縄型産業中核人材育成事業「地理空間ビッグデータ活用『持続可能な観光開発』人材育成プログラム」
一般社団法人環境共創イニシアチブ(SII)	「平成29年度産業データ共有促進事業補助金」に係る「データ活用コンテスト開催」
さくらインターネット株式会社	「平成30年度政府衛星データのオープンアンドフリー化・データ利用環境整備事業」 (第1回・第2回コンテスト・対面講座)
	「平成31年度政府衛星データのオープンアンドフリー化・データ利用環境整備事業」 (第3回コンテスト・e-Learningコンテンツ開発・ラーニングイベント設計)
国立研究開発法人 新エネルギー・産業技術総合開発機構 (NEDO)	「高効率・高速処理を可能とするAIチップ・次世代コンピューティングの技術開発／AIチップ・次世代コンピューティング分野の技術開発の効率的・効果的实施に係る調査」
	「Connected Industries推進のための協調領域データ共有・AIシステム開発促進事業 AI Quest (課題解決型AI人材育成事業) に関する調査事業」
	「高効率・高速処理を可能とするAIチップ・次世代コンピューティングの技術開発／高効率・高速処理を可能とするAIチップ・次世代コンピューティングの技術開発に係るアイデア発掘のための課題調査」
国立研究開発法人 海洋研究開発機構(JAMSTEC)	海洋地球ビッグデータを活用したAIコンペティション企画・運営業務
広島県	令和元年度ひろしまサンドボックス AI人材育成チャレンジ実施業務
国立国会図書館	OCR処理のためのレイアウト認識プログラム研究開発作業

<https://signate.jp/>

3. AIコンペの企画案 (コンペの課題、審査基準、スケジュール等)

実施背景

AI-OCRによりデータセット化した貿易統計情報を活用した戦略立案AIコンペとして、農林水産物の輸出額のポテンシャルを競うコンペティションを企画

タイトル

第1回 農林水産省データサイエンティストチャレンジ ～農林水産物・加工品の輸出ポテンシャルを可視化せよ！～

背景

社会環境

- ✓ 少子高齢化・人口減少に伴い国内における食市場は縮小する可能性が大きい
- ✓ 一方、世界においては人口の増加・経済発展により食市場が急速に拡大しており、日本の農業・食品業としては、「なにを」「どれだけ」「どこ向けに」生産・輸出するかの輸出戦略の重要性が増している

政策立案手法

- ✓ 政策立案においては、限られた資源を有効に活用し、国民により信頼される行政を展開するために、従来の「事例・体験」、「論理的正しさ」だけでなく、データを適切に分析して客観的に意思決定すること(EBPM)が重要となっている
- ✓ また、EBPMの推進においては、データの取得や分析に機械学習等のAIをいかに活用していくかの具体化を早急に実施することが求められている

コンテスト設計

成長市場と成熟市場では、戦略は大きく異なるとの仮説より、これらを別部門として2部門制でコンペティションを設計

部門	成長市場部門	成熟市場部門
課題	✓ 対象国はインド、インドネシア、ミャンマー、エチオピア等の成長市場 ✓ 対象国における5年後の輸出ポテンシャル及びそのポテンシャルを実現する施策の仮説を期待する	✓ 対象国はアメリカ、EU、イギリス、ロシア等の成熟市場 ✓ 対象国における現時点の輸出ポテンシャル及びそのポテンシャルを実現する施策の仮説を期待する
審査基準/提出物/ルール	✓ 参加者は、品目・輸出国を1つずつ選択のうえ、輸出ポテンシャルについて1つ以上の仮説を立てて、貿易統計データ等を用いて定量的な分析を行う。参加者は、以下の3つを提出する ✓ 選択した品目・輸出国における輸出ポテンシャル(億円) ✓ 仮説と、それに基づいて作られた分析アルゴリズム ✓ 立てた仮説の説得力を補完する資料(貿易統計データの定量分析含む) ✓ 複数の審査者は各仮説の内容を検証・検討し、実現可能性スコアを0%から100%の間で決める ✓ 仮説は、輸出ポテンシャル(億円)×実現可能性スコア(%)で算出される実効輸出ポテンシャル(億円)で順位付けがなされる ✓ 実現可能性スコア(%)は以下の2つの視点より検証・検討を行う ✓ 輸出国における品目の現在の市場規模についての妥当性(日本の輸出以外も含む) ✓ 輸出国における品目の需要拡大あるいは日本から輸出品目のシェア拡大を図る施策の効果及び実現性 ※日本の生産・輸出能力には、実効輸出ポテンシャルを踏まえ、農林水産省が強化を図る領域であるため制約を設けないこととする	

成熟市場部門

対象国 : アメリカ、EU、イギリス、ロシア等、すでに市場が安定成熟している国

テーマ : 現在(2020年)の日本の輸出ポテンシャルの試算と市場形成のための施策立案

部門説明

成熟市場部門のゴールは、既に安定している巨大市場に対して、データに裏付けられた輸出戦略を立案することです。

経済的に発展を遂げた国では、既に海外から多数の食品を輸入しています。自国内での生産能力も高く、一般に市場競争は熾烈になります。すでに顕在化されている巨大市場を有する国は、輸出対象として当然重要と考えられます。

そのような競争環境下で、各国において国内生産されている品目あるいは、日本以外からの輸入が多い品目などをリプレイスすることは可能でしょうか？また、成熟市場においても、まだ取引の少ない新しい品目の輸出量を増やすことは可能でしょうか？

実は、我々は成熟市場に対する適切なアプローチに気づいていないだけで、機会損失をしているのかもしれない。

輸出事例として、うまくいった例を紹介しましょう。北海道の十勝地方は「ながいも」の一大産地です。しかし、豊作時に価格が暴落してしまったり、大き過ぎるも(太物)のはカットされてしまい低価格でしか販売できないという課題がありました。そんな中、台湾では、医学界が「ながいも」を漢方薬と位置付けたことをきっかけに、薬膳料理のスープ食材として人気になりました。この薬膳ブームに着目した仲卸業者が、「十勝川西長いも」を台湾に輸出しました。日本では好まれない太物が台湾では逆に好まれ、台湾産より色が白くおいしいため、高い評価を得ることとなりました。詳しくは[コチラの記事](#)を参照ください。このようなケースは他にもありえるのではないのでしょうか？

様々な角度から仮説を立てていただき、データ収集・分析による仮説のサポートを期待します。

成長市場部門

対象国 : インド、インドネシア、ミャンマー、エチオピア等、これから市場成長が期待される国

テーマ : 5年後(2025年)の日本の輸出ポテンシャルの試算と市場形成のための施策立案

部門説明

成長市場部門のゴールは、これから顕在化する市場に対して、データに裏付けられた輸出戦略を立案することです。

今後、人口が増えていき、若者の占める割合も大きい国。インフラや都市開発が進み、暮らしがどんどん豊かになっていく。今はまだ市場が顕在化されていなくても、そんな経済成長が見込まれる国は、この先の輸出対象としてとても魅力的です。かつての日本も高度経済成長期を経て、海外の様々な文化の影響を受け、ライフスタイルも大きく変化し、食生活は多様になりました。現在、私たちは日本にいながらにして、世界の味覚を楽しむことができます。

そのような社会変化に伴う食の多様化に何かしらの法則はないのでしょうか？ 自国と類似した国(類似の定義もいろいろ考えられます)で普及した食品が、その国ではまだ普及していないといったことはあるのでしょうか？そして、何がきっかけで、海外の食品が国内で普及するのでしょうか？

社会の情報化が進む中、世界の情報格差はどんどん少なくなっています。思いもよらないことが、きっかけになるかもしれません。

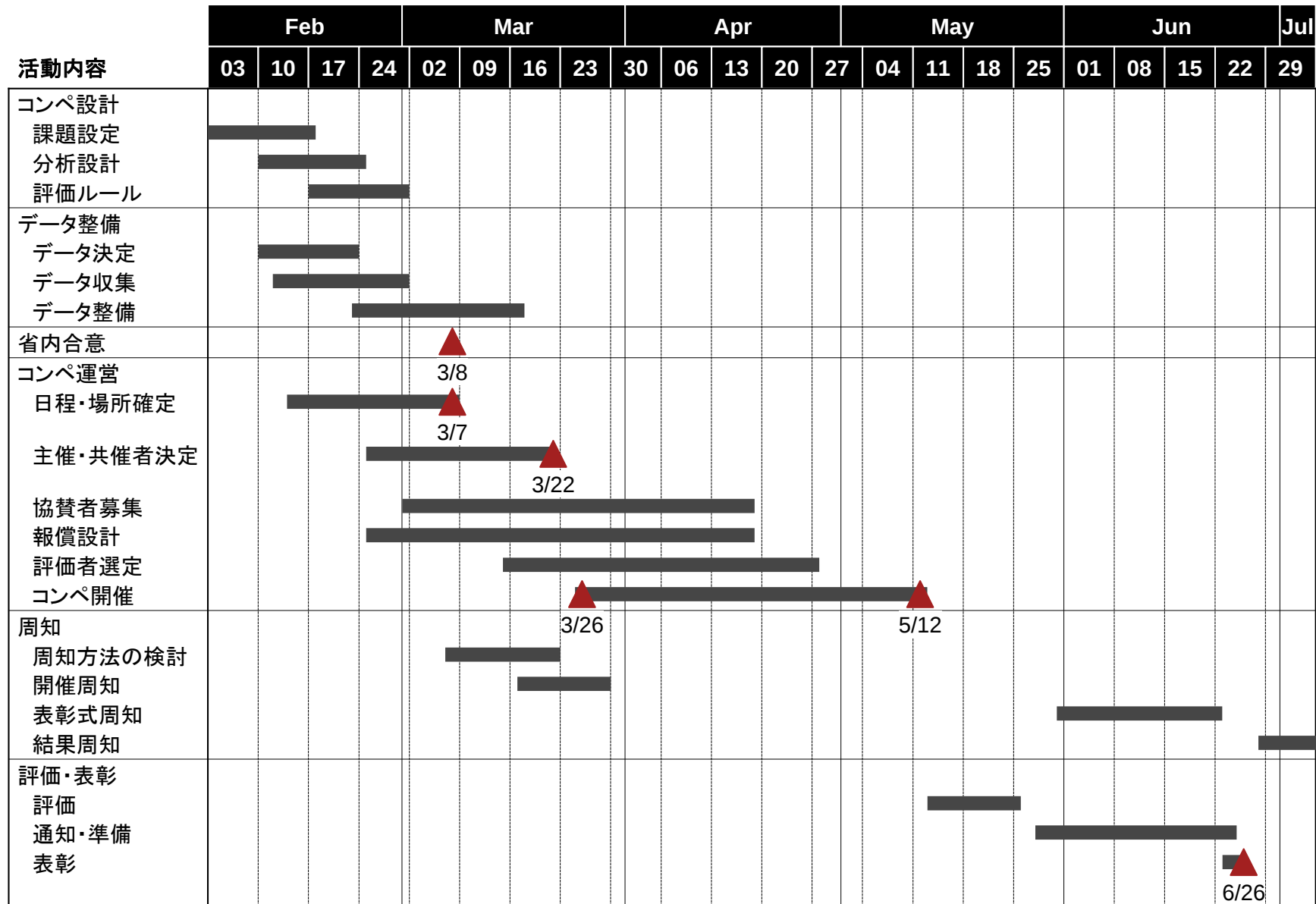
様々な角度から仮説を立てていただき、データ分析による仮説のサポートを期待します。

実施体制・報償

農林水産省を主催者としたうえで、輸出戦略や自社保有データの活用・価値に課題を抱える企業等をスポンサーに加え、コンペ参加者・スポンサーともにメリットを享受できるようにする

役割	企業・団体	報償
主催者	✓ 農林水産省	共催者・協賛者で下記を用意 ✓ 賞金目標:100万円 ✓ その他:食料品等の現物支給 報償イメージ ✓ 部門賞 (成長市場部門、成熟市場部門) ✓ 各部門上位3名に総額100万円 ←共催者・協賛者 ✓ 農林水産大臣or次官との対談+XX室での記念撮影 ←農林水産省 ✓ 協賛企業賞 ✓ 農産物・食品 ←食品メーカー・商社・自治体 イメージ ・トマト賞:トマト1年分(カゴメ) ・メロン賞:メロン1年分(茨城県鉾田市) ✓ 農泊体験 ←自治体 ✓ その他
共催者 (企画運営)	✓ AgVentureLab ✓ SIGNATE ✓ PwC	
協賛者 (スポンサー)	✓ 食品メーカー ✓ 食品商社 ✓ 地方自治体	
データ提供者 (スポンサー)	✓ 日本農業新聞? ✓ Weather News? ✓ ..	
後援者 (周知等)	✓ 経済産業省、他省庁 ✓ 農業・食品関連団体 (メディア含む) ✓ データサイエンス関連団体など	

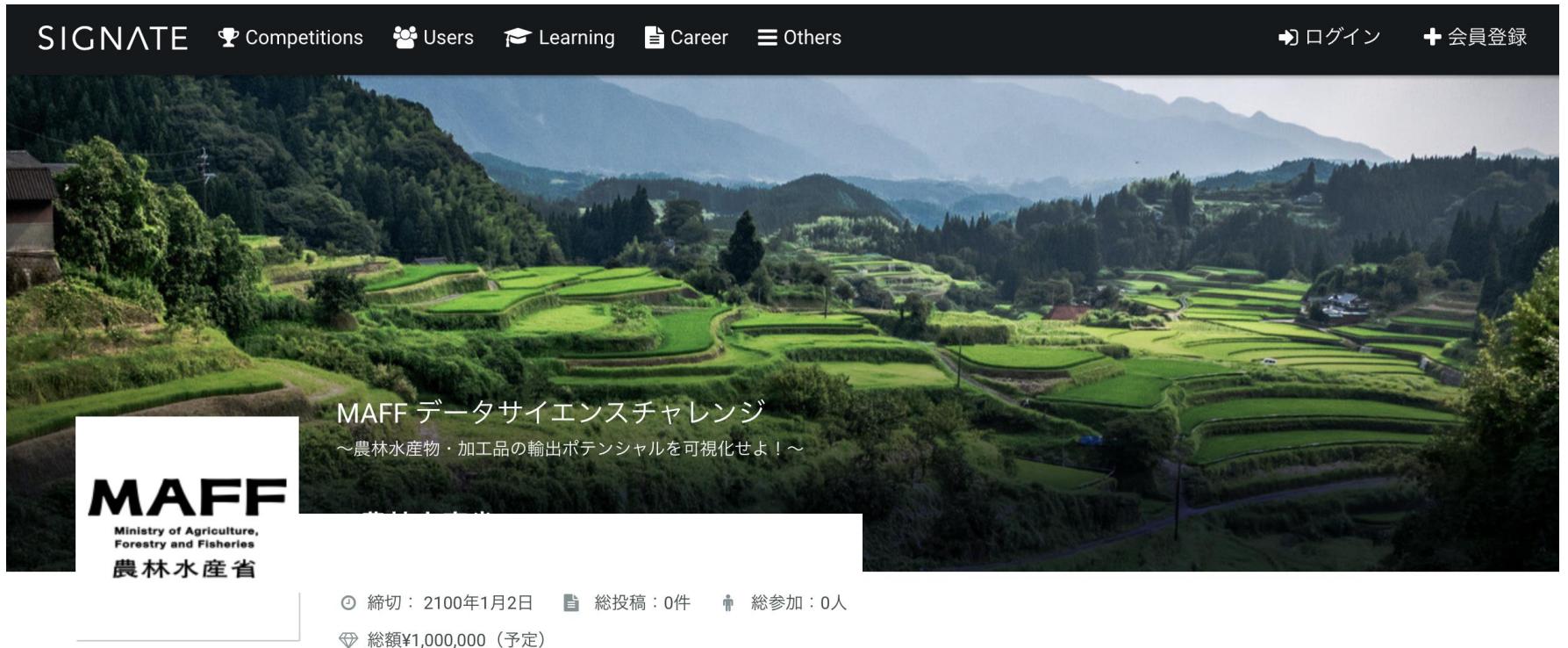
コンペ開催に向けたマスタースケジュール



4. AIコンペ募集サイト

Topページ

28,000人の会員を持つSIGNATE社のコンペティションサイト内で参加者を募る



The screenshot shows the SIGNATE website interface. The top navigation bar includes the SIGNATE logo, links for Competitions, Users, Learning, Career, and Others, along with buttons for ログイン (Login) and 会員登録 (Member Registration). The main banner features a scenic view of terraced rice fields in a mountainous region. Overlaid on the banner is the text for the MAFF Data Science Challenge, including the MAFF logo and the Ministry of Agriculture, Forestry and Fisheries name in English and Japanese. Below the banner, competition statistics are displayed: 締切 (Deadline): 2100年1月2日, 総投稿 (Total Submissions): 0件, 総参加 (Total Participants): 0人, and 総額 (Total Amount): ¥1,000,000 (予定).

SIGNATE ログイン 会員登録

Competitions Users Learning Career Others

MAFF データサイエンスチャレンジ
～農林水産物・加工品の輸出ポテンシャルを可視化せよ！～

MAFF
Ministry of Agriculture,
Forestry and Fisheries
農林水産省

🕒 締切：2100年1月2日 📄 総投稿：0件 👤 総参加：0人
💎 総額¥1,000,000（予定）

https://signate.jp/competitions/257?client_key=d8Dk276qdywYLxAa#abstract

ランキングボード

コンペティションの最初から締め切りまで、コンペ参加者の勝ちたい気持ち刺激し、更なるアイデアを引き出すために、品目・国ごとの輸出ポテンシャルが高い順に暫定ランキングを表示する
今回のコンペ特別仕様のランキングボードを用意

順位	チーム名 / ユーザ名	KPI	国	品目	投稿日時	レポート
1	MTL 	500	インド	ミカン	2019-11-29 00:10:02	
2	phalanx 	300	ブラジル	サケ	2019-11-30 20:13:02	
3	Kocha 	200	カナダ	ミカン	2019-11-29 06:16:02	
4	aoyamalab 	100	ブラジル	リンゴ	2019-11-29 15:32:01	
5	FT 	50	オーストラリア	バナナ	2019-11-29 17:28:02	

フォーラム

フォーラムを用意することで、コンペ参加者からの質問受付・回答など、会員のコミュニケーションを可能とすることで、参加者の理解促進を図り、より良いアイデアを引き出す

説明

データ

ランキング

フォーラム

← FP, FNの扱い

▼ フォーラム機能について



kwea koi

3ヶ月前 (編集済み)

12

評価の際に、FP（誤検出）とFN（未検出）はどういう扱いになるでしょうか。
評価の例にはなかったので聞きたいです。
もしくは、評価スクリプトを公開していただければ助かります。
よろしくお願いします。

1

S

SIGNATE

3ヶ月前

事務局です。

ご質問いただきありがとうございます。
評価関数はIoUなので、誤検出、未検出という概念はありません。
評価方法に関して補足しますと、例えばどの正解領域ともIoUが0となるような予測領域に対するペナルティはなく（そのような領域が1個あっても10個あっても評価値は変わらない）、ある正解領域に対してどの予測領域もIoUが0だった場合はその正解領域に対するIoUは0となりま

データ提供

ログインした会員のみがデータをダウンロード可能とする仕様のため、目的外のデータ取得・活用を抑制する

SIGNATE

CompetitionsUsersLearningCareerOthers

ototo

説明

データ

ランキング

投稿済みファイル

フォーラム

データダウンロード

1

本データは参加規約第4条1項の通り、本コンペティション及び別途当社が指定した目的以外の使用はできません。

ダウンロード	ファイル名	ファイル容量
	学習用画像データ (train_images.zip)	1.32 GB
	学習用画像データのアノテーション (train_annotations.zip)	2.42 MB
	学習用高精細画像データ (hr_train_images.zip)	2 GB
	学習用高精細画像データ (hr_train_images_2.zip)	1.77 GB
	評価用画像データ (test_images.zip)	256.1 MB
	評価用画像データのアノテーション (test_annotations.zip)	232.63 KB
	評価用高精細画像データ (hr_test_images.zip)	812.73 MB
	応募用サンプルファイル (sample_submit.json)	27.96 KB

Image

5. AI/アナリティクスを活用した輸出分析及び 政策立案に有効な情報リスト

AI/アナリティクスを活用した輸出分析及び政策立案の仮説

AI/アナリティクスを活用し輸出分析及び政策立案を行う際の仮説アプローチとして、「品目間の構造俯瞰」、「輸出の増加要因の深掘り」の2つを検討

仮説のアプローチ

アプローチ詳細・仮説イメージ

品目間の構造俯瞰 による仮説

農産品目あるいは他品目の輸出全体構造を俯瞰した仮説

- ✓ 温州みかんを輸入が多い国では、●●の輸入も多い。対象国Aでは温州みかんの輸入に対して●●の輸入が少ないため、拡大ポテンシャルが存在
- ✓ 冷蔵庫の輸入が増加した国では輸入の大幅増加から5年後に●●の輸出が伸びる傾向がある。対象国Bでは冷蔵庫輸入が拡大し始めたので、今後の輸出ポテンシャルが存在。

輸出の増加要因の 深掘りによる仮説

商品特性、輸出国の自然・経済・社会の環境、これら環境とその変化による消費者の嗜好の変化、あるいは、競合品の状態などの特定商品の日本からの輸出に影響を与える因子を深掘りした仮説

- ✓ 日本文化（漫画・アニメ）などの浸透度が高い国では、対象層が成人する段階で●●の輸出が増加傾向にある。対象国Cでは日本文化が浸透し始めているため、今後の輸出ポテンシャルが存在。
- ✓ ●●の輸出においては、対象国の気候区分及び1人当たりGDPが大きく影響している。対象国D国は気候区分として条件に合致しており、今後数年で1人当たりGDPも輸出が増加する基準まで到達するため、今後の輸出ポテンシャルが存在

AI/アナリティクスを活用した輸出分析及び政策立案に有効な情報リスト

前述の仮説のアプローチに基づき、輸出増減因子の分析やポテンシャル推計に有効な情報として、経済、国、地理・気候、貿易等の情報をリストアップ

#	データ	ソース	URL	備考
1	物価指数(日米)	IMF World Economic Outlook Database	https://www.imf.org/external/pubs/ft/weo/2019/02/weodata/index.aspx	
2	為替相場(日米欧)	Euromonitor	https://www.euromonitor.com/	
3	日本の海外旅行者数In/out(総数・国別)、定住者数	JNTO月別・年別統計データ(訪日外国人・出国日本人)	https://www.jnto.go.jp/jpn/statistics/visitor_trends/index.html	
4	国別人口	United Nations World Population Prospects	https://population.un.org/wpp/Download/Standard/Population/	
5	国別1人当たりGDP	IMF World Economic Outlook Database	https://www.imf.org/external/pubs/ft/weo/2019/02/weodata/index.aspx	
6	各国首都の緯度・経度	グーグルアース4	http://www.xn--4-eeuvba0c6t4dd.jp/data/list00.html	
		地名緯度経度データ	http://gpscycling.net/land/latlon/	
7	主要都市の平均気温(月別)	気象庁	https://www.data.jma.go.jp/gmd/cpd/monitor/main/stn/nrmlist.php https://www.data.jma.go.jp/gmd/cpd/monitor/climateview/frame.php	
8	ケッペン気候区分	WORLD MAPS OF KÖPPEN-GEIGER CLIMATE CLASSIFICATION	http://koeppen-geiger.vu-wien.ac.at/applications.htm	リストではなくMap
		(v). Climate Classification and Climatic Regions of the World	http://www.physicalgeography.net/fundamentals/7v.html	
9	国別食品輸出入額	Global Trade Atlas	https://www.jetro.go.jp/db_corner/gta.html	JETROライブラリー
10	国別エネルギー輸出入額	IEA	https://www.iea.org/data-and-statistics?country=WORLD&fuel=Imports%2Fexports&indicator=Coal%20imports%20vs.%20exports	国別項目別時系列は難しい

pwc.com/jp

© 2019 PwC. All rights reserved.

PwC refers to the PwC network member firms and/or their specified subsidiaries in Japan, and may sometimes refer to the PwC network. Each of such firms and subsidiaries is a separate legal entity. Please see www.pwc.com/structure for further details.

This content is for general information purposes only, and should not be used as a substitute for consultation with professional advisors.