




**企画番号：24**

**企画タイトル：AI やデータ分析を活用し、実際に  
役立つものを作る**



# プロジェクトリサーチ 活動報告書

## No.24 AI やデータ分析を活用し

## 実際に役立つものを作る

代表：Y200042 小橋口 純  
副代表：Y200065 玉井 信  
Y200004 安部 凌平  
Y200051 猿渡 脩大  
Y200062 竹本 来生  
Y200078 前川 悠人  
Y200080 松本 虎太郎  
Y200089 森 昂誠  
Y200090 森 裕都

# 概要

## 目的：

数理・情報科学課程の学生の強みは、技術の仕組み、構造などを理解した上で活用し取り扱うことができることである。この強みを生かし、進歩し続ける A I（人工知能）などの情報技術を活用し、実際に役立つものを作りたいと考えた。

今回の活動では、情報技術を生かし、学習進度や学習目的に応じて自動的に学習内容を調節する機能を持つ、英単語学習支援サービスを作成することを目的とした。

## 計画：

具体的な活動への円滑な移行のため事前に、既存のサービスや取り組み、人工知能、データ分析、統計学について調査・学習を行なう。上記の項目を学習後、具体的な取り組みとしてはブラウザ上で動く英単語学習アプリを作成し、班員によるデータ収集から得られた結果を基に個人個人のデータ分析を行なう。その後、結果を活用し学習効率の向上を目的とした個別最適化された学習支援サービス構築を目指す。

## 方法：

Web アプリ作成の方法について、英単語学習、アプリの開発方法、アプリの形態、デプロイ環境を事前調査し、Python の Web アプリケーションフレームワーク「Flask」と Microsoft Azure の「App Service」を使用し、Web アプリを作成する。データベースには軽量で高速に動作する SQLite3 を使用した。

## 活動経過：

時期	内容
6 月	Web アプリを作成することが決定
7 月	作成方法決定
8 月	出題アルゴリズム決定・データ分析の勉強
9 月	アプリのデプロイ
10 月	アプリ完成・データ収集・ポスター、報告書作成

## 成果・結果：

ブラウザにアクセスすることで、誰でも実際に試すことが出来る学習アプリを作成することが出来たが、そのデータを用いた詳しい分析を行うまでには至らなかった。

今回は英単語を表示しその日本語訳を 4 択で回答する形式に限って実装した結果、全体的に正答率が高くなりすぎてしまったと考えている。出題アルゴリズムの修正や、日本語訳を表示しその意味を持つ英単語を入力する形式も実装することで、より効率よく、快適に学習及びデータ収集を行えるようになるのではないだろうか考える。

# 報告書

## 目的：

私たち数理・情報科学課程の学生の強みは、技術の仕組み、構造などを理解した上で活用し取り扱うことができることである。この強みを生かし、昨今個人の力として必要とされている、英語力を向上させるものを作成していく。具体的な構想としては、英語に関する学習アプリの制作を行い、ユーザーの正答率などの情報を元にデータ分析を行なう。その後、AI がユーザーの苦手分野を解明しその分野のみを集めたテストの作成、改善方法の提示などのユーザーごとへの最適化されたものを作る。

このような学習アプリを作成することで、検定や日常会話、専門分野への活用など様々な用途に合わせることで個々人に満足してもらうことを目指す。

## 計画：

- 6月上旬 : Web 上で動く英単語学習アプリを作成することが決定。
  
- 6月中旬 : 参考のために既存のサービスや取り組みについて学習
  - ・英文のデータを的 AI に分析
  
- 6月下旬～7月上旬 : アプリの作成方法の決定、作成開始  
英単語データの引用元の決定
  
- 7月中旬～7月下旬 : 人工知能、データ分析、統計学の学習
  - ・データ分析、統計学の学習  
→学習成果の分析から個別最適化された学習案の作成
  - ・人工知能の学習  
→学習者のデータから今後の学習予測
  
- 8月上旬～8月中旬 : 出題アルゴリズム決定、アプリのデプロイ・データ分析の勉強
  - ・単語学習→確認テストのアルゴリズムの決定、作成開始
  - ・Azure を利用してアプリのデプロイ方法の勉強
  - ・班員から収集したデータの分析方法の勉強
  
- 8月下旬～9月上旬 : アプリ完成、データ集め開始
  - ・アプリの動作確認、個別学習案作成のため班員のアプリの 試行
  
- 9月中旬 : ポスター、報告書作成

## 方法：

Web アプリ作成の方法について、事前に、英単語学習、アプリの開発方法、アプリの形態、デプロイ環境を調査した。

サーバーサイド（バックエンド）には、Python の Web アプリケーションフレームワークである「Flask」を使用した。Python の Web アプリケーションフレームワークは、主に「Flask」と「Django」の二つがある。Django は、Web 開発に必要な機能が初めから盛り込まれている重量級フレームワークという特徴があり、Django だけで基本的な Web アプリケーション開発は完結する。そのため、大人数が関わる大規模な開発に向いている。それに対して、Flask は軽量でかつシンプルという特徴があり、最低限の機能を備え、必要に応じて機能を拡張するという設計思想から初心者にも理解しやすく小・中規模の開発に向いている(1)。以上の理由から今回は学習コストが低く機能が絞られている Flask を選んだ。

クライアントサイド（フロントエンド）には、Web ページを作成するために必要なマークアップ言語である「HTML」、Web ページのデザインを変更するスタイルシート言語である「CSS」、動的に Web ページを作成できるコンピュータ言語「JavaScript」を使用した。

次に、データベース管理システムに利用した SQLite について、ファイルベースでデータを管理するリレーショナルなデータ管理システムであり、軽量で動作が早い特徴がある。

Web サーバーにデプロイする際は Azure の「App Service」を利用した。OS は「Windows」と「Linux」系から選択ができ、開発言語も「.NET」「Java」「Ruby」「Node.js」「Python」など様々な言語から選択できることに加え、Azure の利点である「負荷分散」「自動スケーリング」等のメリットを享受することができる(2)。また、GitHub リポジトリから継続的なデプロイが可能のため、ユーザーは開発に注力できる魅力がある。

## 活動経過：

日付	内容
6月16日	英単語アプリを作成することが決定
6月中旬～	Python の勉強
6月中旬～	HTML の勉強
6月中旬～	CSS の勉強
6月中旬～	JavaScript の勉強
6月中旬～	Flask の勉強
7月3日	レイアウト、デザインの考案
8月上旬～	データ分析の勉強
8月上旬～	Azure の勉強
9月4日	単語データの分類プログラム作成
9月6日	出題アルゴリズムの決定
9月15日	出題アルゴリズムの改良
9月17日	問題出題プログラムを Web アプリに移植
9月24日	英単語アプリのデプロイ
10月10日	英単語アプリの完成
10月13日～10月24日	データ収集
10月24日	データ分析
10月24日～10月27日	ポスター作成
10月31日～11月4日	報告書作成

### <主な担当業務>

小橋口 純 : データ分析、前半のスケジュールマネジメント、Azure へのデプロイ

玉井 信 : 出題アルゴリズムの構築、API の開発

安部 凌平 : データ収集、Web アプリの動作テスト&フィードバック

猿渡 脩大 : データ収集、フロントエンド言語の勉強

竹本 来生 : データ収集、書記、データ分析

前川 悠人 : データ収集、Web アプリの動作テスト&フィードバック

松本 虎太郎 : Web アプリの一部機能開発、後半のスケジュールマネジメント

森 昂誠 : Azure の勉強、Web アプリの動作テスト&フィードバック

森 裕都 : Web アプリの全体開発、Azure へのデプロイ

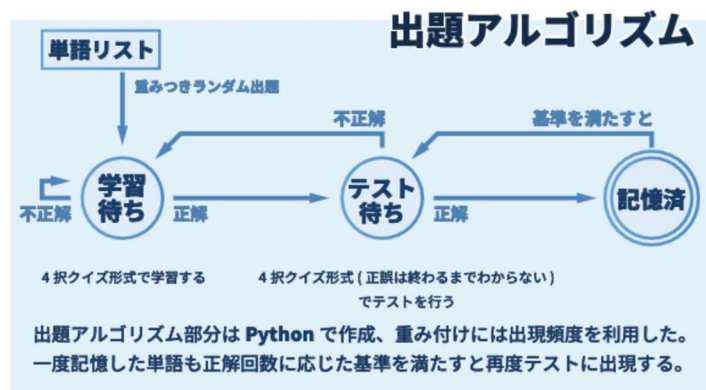
## 成果・結果 (Web アプリ) :

### ・Python を使った出題アルゴリズム

プログラムを組む準備段階として、まずは各単語に重複しない ID を割り振った。この ID は CEFR ランク、品詞、通し番号からなる 8 桁の自然数で、絞り込み検索や逆引きなどに活用できるようにこの形式とした。実際、配列等に読み込んだ際の要素番号と一致させた下 5 桁の通し番号は、プログラムを組むうえで大いに役立った。

出題アルゴリズムは以下のように設定した。

1. 学習する単語を難易度別単語リストから重み付きランダム抽選
2. クイズで正解するとテスト待ち状態へ遷移、不正解だとそのまま
3. テスト待ち状態の単語が一定数貯まるとテスト受験可能に



4. テストで正解すると記憶判定、不正解だとクイズへ再度遷移する
5. 記憶判定の単語も暫く経つと復習待ちとなり、再びテストに出現する

重み付けには、単語ごとの使用頻度を利用。on や the などの日常的に多用する単語は、あまり使用しない単語と比べて約 3 倍出現しやすい設定にした。

復習待ちへの遷移条件には総学習回数を利用。ある単語の学習後に約 50 単語学習すると、ある単語は復習待ちへと遷移する。この遷移までに必要な回数は、その単語を連続正解するごとに増えていく。そのため、簡単な単語は連続正解することで徐々に出現しなくなり、逆に難しい単語は復習の段階で間違えるなどして再度出現しやすくなっている。

### ・Web アプリの概要

ユーザーごとに成績情報を残すため、ユーザー認証機能を実装した。ユーザー名やパスワードなどのユーザー情報は SQLite に保存され、クイズやテストの結果が随時記録・更新されていく仕様になっている。

ログイン後、ユーザーはマイページから、

- ・成績やクイズの進捗状況などをパラメータやグラフで確認できる「ダッシュボードページ」
- ・Web アプリに登録されている全英単語の情報を確認・検索できる「ライブラリページ」
- ・ランクごとに分けられたクイズに挑戦できる「学習ページ」

- ・クイズで正解した英単語の確認テストを行える「確認テストページ」
- ・アプリ使用上の疑問点などを解決できる「FAQ ページ」
- ・ユーザー情報の変更やダークモードの切り替え、ログアウトなどを行える「設定ページ」

を利用することができる。

以下実際の Web アプリの画像



【ホーム】



【ダッシュボード】



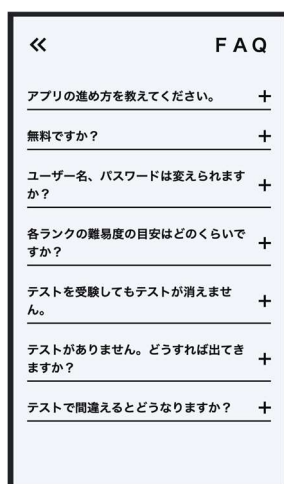
【ライブラリ】



【クイズ】



【テスト】



【FAQ】



【設定】



## 成果・結果（データ分析）：

Web アプリでは、SQLite を用いて学習のデータを記録している。そのデータを用いたデータ分析にも挑戦した。データ分析の目的は2つ設定した。1つ目は、Web アプリに学習効果があるのかどうかを判断することである。2つ目は、覚えやすい／覚えにくい英単語の特徴を考察し、出題アルゴリズムの改善に生かすことである。

アプリの完成後各メンバーは、2週間にわたり、アプリを利用して英単語学習に取り組み、データを作成した。アプリのダッシュボード機能を用いて学習した単語数を把握し、200語程度学習した時点でデータ作成を終了した。その後、SQLite に保存したデータを csv ファイルとして出力し、Python ライブラリの Pandas を用いてデータ分析を実施した。

csv ファイルに出力したデータは、先頭ラベルが「英単語」であり、各行が1回分のクイズ・テストに対応する。正解は1、不正解は0、出題されなかった場合は-1を記録するようにした。Pandas を用いて、-1を np.nan に置き換えた。

しかし、データが少なかつたため、学習効果があるかどうか判断することはできなかつた。学習した単語の総数としては200語程度だったが、一つ一つの単語については1～10回程度しか出題されていなかつた。そのため、学習者が各単語を本当に覚えたかどうか判定することができなかつた。

出題回数が少なくなつた原因は、出題アルゴリズムにある。まだ覚えていない単語が多くあつたとしても、「クイズ」を開始することにより、無制限に新しい単語の学習を始めることができる。データ分析が可能なデータを集めるという観点からも、「覚えたという判定」を出すようにし、それに伴って新しい単語の学習を開始できるようにすべきだと考えられる。

また、覚えやすい／覚えにくい英単語の特徴を分析するために、英単語の特徴データを作成した。その際は、品詞の one-hot 符号化やデータの正規化を行った。その上で、各英単語の正答率を目的変数、英単語の特徴量（文字数、出現頻度、品詞）を説明変数として重回帰分析を行った。しかし、決定係数が低く有効な分析はできなかつた。

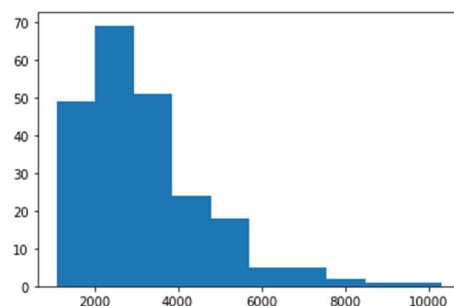
その他に、回答時間の分布についてのヒストグラムを作成した。

	abnormally	abolish	abruptly	absentee	al
0	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...
140	NaN	NaN	NaN	NaN	NaN
141	NaN	NaN	NaN	NaN	NaN
142	NaN	NaN	NaN	NaN	NaN
143	NaN	NaN	0.0	NaN	NaN
144	NaN	NaN	NaN	NaN	NaN

145 rows × 1822 columns

	abnormally	abolish	abruptly	absentee
frequency	0.000085	0.000706	0.000291	0.000138
adjective	0.000000	0.000000	0.000000	0.000000
adverb	1.000000	0.000000	1.000000	0.000000
conjunction	0.000000	0.000000	0.000000	0.000000
determiner	0.000000	0.000000	0.000000	0.000000
interjection	0.000000	0.000000	0.000000	0.000000
noun	0.000000	0.000000	0.000000	1.000000
preposition	0.000000	0.000000	0.000000	0.000000
pronoun	0.000000	0.000000	0.000000	0.000000
verb	0.000000	1.000000	0.000000	0.000000
log_frequency	0.185917	0.362181	0.286198	0.224270
length	0.571429	0.357143	0.428571	0.428571
category	NaN	NaN	NaN	NaN

13 rows × 1822 columns



## 今後の展望：

### ・学習する単語セットに関する改善

アプリ内で採用されている英単語のリストは、英語をはじめとした外国語の習熟度や運用能力を同一の基準で評価する国際標準である CEFR を元に製作されたものを採用したが、他にも高校受験、大学受験、プログラミングなど様々な環境に特化した英単語のリストを採用することで、他の英単語学習アプリと差別化を測れるのではないかと考えている。

### ・アプリ内の学習アルゴリズムに関する改善

現状のアルゴリズムにおいて、記憶判定された単語が復習待ちに帰って来るまでの期間は  $50 \times (\text{その単語の連続正解回数})^2$  回だけテストで正解したらという風に設定されている。しかしこの基準に関してはまだ改善の余地があると考えており、エビングハウスの忘却曲線などの科学的根拠に基づいて再構築することで、より効率の良い学習体験を提供できるのではないかと考えている。

### ・アプリの機能に関する改善

各ユーザーの成績データを見ると、そのユーザーが苦手としている単語を抽出することができるので、それを元にそのユーザーにパーソナライズ化された小テストを生成する機能を増設したいと考えている。

### ・データの収集に関する改善

今回の活動期間では十分なほどデータの収集が出来たとは言えないため、今後もデータ収集を続けていくことで、例えば多くの人の間違えやすい英単語の特徴など、現状では見つけ出せなかった何かしらの英単語の共通点を見つけ出すことが可能になるかもしれない。また、データベースの構造を見直して収集できるデータの種類を改善することで、より良いデータ収集・分析に繋がられるのではないかと考えている。

## 感想・学んだこと：

今回の活動は自分達が主体となって調べ、行動することで進めていった。これから始まる卒業研究も、同じように自らが主体となって取り組んでいかなければならないものなので、この経験を活かしていきたいと考えている。(松本 虎太郎)

データ分析を成功させるには、目的を明確にし、どのようなデータを集めるか決めてからデータ収集を行うことが重要である。データ分析の計画を踏まえてアプリを設計することも必要だと分かった。

今回の活動では、プログラミング、データ分析、デザイン、教育、チームマネジメントなど、メンバーそれぞれの得意なことや興味のあることを生かして活動を進めることができた。

(小橋口 純)

それぞれの予定もあり計画通りに進めることが難しく、目標としていたことが出来なかった。任せきりになってしまった部分もあったため役割分担の必要性を感じた。

メンバーと話し合い改善することでアプリが完成していくといった過程が楽しく良い経験ができた。(猿渡 脩大)

今回自分たちで Web アプリを作る際、利用者にどうすれば効率よく単語を覚えられるのかについてよく議論をしていました。議論の結果できたアプリにはまだまだ改善点は多かったが、アプリをよくするためにいろんな意見を出し合って努力したことは、プロジェクトリサーチをやって良かったことだと思います。プロジェクトリサーチで経験したいろんな意見を出し合って改善をしていくことが、社会人となって多くの人と仕事をする際、大いに役に立つと思います。(前川 悠人)

Web アプリの作成からサーバーへのデプロイに至るまで、全ての行程に携われたことは良い経験になった。今まで全く使ったことのない言語を一から勉強していくことは大変ではあったが、今回の活動を通して、新しい技術に挑戦することの面白さを実感できた。今後は、少人数でのチーム開発、他の言語や Azure 以外のサービスを活用した Web 開発に取り組んでいきたい。(森 裕都)

ゼロの部分から、発足、計画、考案などを経て1つのプロジェクトを立ち上げ、それを形にするこの大変さを痛感しました。多くの人数が参加しているからこそ、時折メンバー間でのモチベーションの差が生じてしまうことの難しさなどを感じました。社会に出るとこれよりも多くの人に関わる事もあるのでプロジェクトを動かすために重要なことを学ぶことができたと思う。

(安部 凌平)

Python で本格的にアルゴリズムを組んだのはこのプロジェクトが初めてだったため、とても勉強になった。この経験を生かして、作業の効率化などに活用できるようなプログラムを今後も作成していきたい。(玉井 信)

英単語アプリの作成からデータ分析まで、数理・情報科学課程ならではの数学と情報、両方の知識を活かし取り組むことが出来た。データ分析が主な担当ではあったが、分析するための情報量がどの程度必要か何に対しての分析が重要なかなど難しさがあった。今後、卒業研究に打ち込んでいくうえで、今回の計画から展示会までのプロセスを見つめなおし、活かせるようにする。(竹本 来生)

A I や今までの学部での学びを活かして英単語アプリを作成できたのは非常に良い経験になった。実際にプログラミングなどで大きな力にはなれなかったが、プログラミングに関しての学習意欲が芽生えたし、今後のチーム活動の際に大事になることも今回の活動を通して得られたと感じるので活かしていきたい。(森 昂誠)

#### Web アプリ作成にあたり参考にしたもの：

- ・ CEFR-J Wordlist  
<http://www.cefr-j.org/download.html>
- ・ ANC 単語頻度準拠\_英和辞典  
<https://www.jamsystem.com/ancdic/index.html>

#### 報告書の参考文献：

- (1) 「【徹底比較！】 Flask と Django それぞれの特徴と選定基準について」 ENGINEER LIFE (2022 年 11 月 4 日確認)  
<https://engineer-life.dev/flask-django/#:~:text=%E3%81%A8%E6%80%9D%E3%81%84%E3%81%BE%E3%81%99%E3%80%82-,Flask%E3%81%AE%E7%89%B9%E5%BE%B4,%E3%82%B7%E3%83%B3%E3%83%97%E3%83%AB%E3%81%AA%E3%83%9E%E3%82%A4%E3%82%AF%E3%83%AD%E3%83%95%E3%83%AC%E3%83%BC%E3%83%A0%E3%83%AF%E3%83%BC%E3%82%AF%E3%80%82>
- (2) 「Azure App Service とは？概要と基本的な使い方をご紹介」 Rainbow Engine (2022 年 11 月 4 日確認)  
<https://rainbow-engine.com/azure-app-service-overview/>