

# 学生生協における AI を用いたマーケティングの利用

1200545 横田 和哉

高知工科大学 経済・マネジメント学群

## 1. 概要

日常生活・産業界でも利用が進んでいる AI を用いマーケティングを行うこと学生生協における機会損失、廃棄ロス、自従業員負担の削減を目指すのが本研究の目的である。

特徴量やハイパーパラメータの調整を行うことで、AI の予測精度を向上されることができたが、平均より外れた値の予測は難しく、外れ値の排除若しくは外れ値を捉える特徴量の設定が今後の課題になる。

## 2. 背景

### 2-1 AI の発展

AI の発展は凄まじく、2016 年には囲碁の世界では当時世界最強とされていたイ・セドル九段への勝利や、自動運転技術の発達など我々の産業・生活にまで AI は利活用され始めている。日本の企業でも主要企業の 120 社の内 7 割ほどが AI の導入と AI の導入を検討している。また AI はビッグデータを比較的取得しやすいマーケティング分野との相性がよく、web 上でのターゲティング広告や、小売店では需要予測にも活用する企業が出てきている。

### 2-2 大学生協でのマーケティングの現状

大学生協の売店では、主な顧客が大学の学生であるため、通常の小売店舗とは性質が少し異なる。

通常の小売店との相違点として、以下の点が存在する。

- ・メインのターゲット層が大学生及び職員である
- ・大学の授業の有無・イベントなどにより来客が大きく左右され、1 年の 3 分の 2 ほどが休みになる
- ・陳列商品・価格の決定は店舗の上部組織である事業連合の決定に従う
- ・バックヤードが持たず、小さな大学の生協ではマーケティングに人を裂ける余裕がない
- ・学生生協の存在目的は「学生の学びを支える」

本研究で対象とした高知工科大学永国寺キャンパスの大学生協でも上記にあげた性質を持ち合わせており

- ・マーケティングに人と時間をさけない
  - ・バックヤードを持たないため毎日従業員が勘と経験に基づいて発注を行っている
- というマーケティング上の課題を抱えている。

この大学生協でのマーケティング上の課題を、AI を用いて解決することができないかと筆者は考え、研究を進めることとした。

## 3. 研究目的

本研究では、大学生協において AI を用いてマーケティングを行うことで、生協での廃棄ロス・機会損失の減少、従業員の負担の減少を目的としている。また、マーケティング上起こしうるバイアスや留意点、学生生協ならではのマーケティング特性の発見も目的とする。

## 4. 研究方法

### 4-1 概要

本研究では、商品売上予測（1 ヶ月毎）とそのために必要な来店客数予測（日毎）の二つのモデルを、ディープニューラルネットワークを用いて作成する。学習・検証は研究室内の解析ソフトを用いて行う。モデルはシーケンシャルモデルを用いる。

### 4-2 研究の流れ

- ①提供されたデータに筆者が結果に影響を与えうると考えた特徴量をデータに追加
- ②データから欠損値をもつデータの排除
- ③AI を用いて学習・検証
- ④結果の検証
- ⑤特徴量・ハイパーパラメータ調整
- ⑥②～⑥を繰り返しモデルの精度向上を図る

### 4-3 研究対象

研究対象は高知工科大学永国寺キャンパス生協売店であり

研究に用いるデータは生協売店の従業員が独自に集計していたデータを用いる。また永国寺キャンパスを対象にしているため、学生数や授業の有無などは高知工科大学（以下 工科大）の場合はマネジメント学群及び、経済・マネジメント学群の2～4年生、高知県立大学（以下 県立大）の場合は文学部の1～4年生を対象としている。

#### 4-4 予測対象データ

##### 4-4-1 来店客数予測

来店客数予測の予測対象データは2015年4月～2019年12月までの日毎のデータを用いる。データの中身は対象日、対象日の永国寺キャンパス生協ショップの午前客数（9時～12時）、午後客数（13時～）を含んでいる。また営業していない日はデータから取り除く。

結果変数は日毎の午前・午後の客数である。

基本統計量（午前客数）							
平均	中央値	最頻値	標準偏差	最小	最大	個数	
93.73	100	22	52.23	0	237	1069	

（図1 基本統計量 午前客数）

基本統計量（午後客数）							
平均	中央値	最頻値	標準偏差	最小	最大	個数	
93.3	99	14	55.86	0	327	1069	

（図2 基本統計量 午後客数）

基本統計量から、平均がどちらも93付近である。標準偏差が55とデータのばらつきが大きいことがわかる。

本研究では、来店客数予測では、午前客数の予測は行わず午後客数の予測をおこなう。

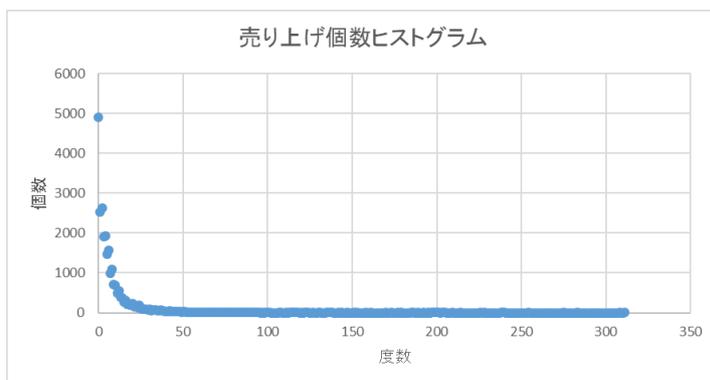
##### 4-4-2 商品売上予測

商品売上予測の予測対象データは2017年1月～2019年12月までの月毎のデータを用いる。データの中身は永国寺キャンパス生協ショップに陳列された、商品名、商品カテゴリー、商品の値段、対象月、対象月の売上を含んでいる。対象の商品は、お菓子類、飲料類、デザート類、ご飯類、パン類である。

結果変数は対象月の対象商品の売上である。

基本統計量（月毎売上）						
平均	中央値	最頻値	標準偏差	最小	最大	個数
8.04	4	0	14.09	0	311	26039

（図3 基本統計量 月毎商品売上）



（図4 売上個数ヒストグラム）

ヒストグラムからわかるように、値が0に集中している（値が0の場合は4905個、値が1の場合は2520個）。これはデータの中で店舗の棚に商品が陳列されなかったため、売れなかった商品を含んでいるからである。

#### 4-5 結果の検証・モデル精度向上

結果の検証はMAE（平均絶対誤差）を用いて行い、値が小さくなるよう特徴量、ハイパーパラメータの調整を行う。

モデルの精度向上の手法は大きくわけて3つの方法で行う。

##### 1. ハイパーパラメータの調整

ノード数、層の厚さ、エポック数の調整を行う。エポック数とはデータに対する学習の回数であり、少なすぎると学習不足、多すぎると過学習を引き起こす。ノードは隠れ層の中に存在し、特徴量及び前の層のノードからAIが新たな特徴量を定義したものである。一般的に多い方が良いとされている。層の厚さは隠れ層の数であり、多すぎると収束しすぎ、少なすぎると発散してしまう。エポック数はlossの値が小さくかつ、訓練データのlossとテストデータのlossの差が小さくなる様な数値を探る。ノード数、層の厚さは適宜調整を行い、MAEが小さくなる値を探る。

##### 2. 多重共線性を持つ特徴量の排除

多重共線性とは説明変数（特徴量）同士が高い相関を持つ

ことで起きる現象で、モデルの予測精度を下げってしまう効果をもつ。特徴量同士の相関を測定し強い相関を持つ特徴量のうち片方を排除する。(相関係数 0.4 以上及び-0.4 未満の場合)

### 3. 予測値と実測値の絶対誤差に関する特徴量の排除

検証データの実測値と予測値の絶対誤差を結果変数、検証データの特徴量を説明変数とし重回帰分析を行う。

上記結果から絶対誤差に正に関係する特徴量を排除する。

## 4-6 特徴量設定

特徴量とは、予測の結果を特徴づけるものあり、モデルの作成において重要な役割を果たす。本研究では最初に多めに特徴量を設定し、モデルの作成・検証・修正の過程で適切な量に減らしていく方法で特徴量を調節していく。また適宜特徴量の変更や追加も行っていく。

### 4-6-1 来店客数予測

予測対象である、来店客数は日毎の午後のデータのため、特徴量も日毎で設定する。

特徴量	備考	数
平均気温	当日	1
天気概要	午前、午後、雨、晴、曇 それぞれ 2 値	6
平均風速	当日	1
平均湿度	当日	1
降水量	当日	1
曜日	月～日で 2 値	7
祝日	2 値	1
年	2015～2019 で 2 値	5
月	1～12 で 2 値	12
授業概要	工科大、県立大ごとに 授業有無、テスト有無、予備日、集中講義をそれぞれ 2 値	8
学生数	工科大、県立大、県立大夜間	3
学生男女比	工科大、県立大、県立大夜間	3
前日来店客	午前、午後	2
前日営業	あり、なしで 2 値	1
合計		52

(表 1 来店客数予測特徴量)

### 4-6-2 商品売上予測

予測対象である商品の売上予測は月毎のデータのため特徴量も月毎で設定する。

特徴量	備考	数
値段	連続値	1
期間限定	2 値	1
甘い	2 値	1
飲み物	2 値	1
商品ジャンル	食べ物は分類コード毎、 飲み物は味、容器サイズ、容器種類 毎に 商品の特徴に当てはまる変数でそれぞれ 2 値	56
年	2017～2019 で 2 値	3
月	1～12 で 2 値	12
授業概要	工科大、県立大ごとに その月の授業日数、テスト日数	4
学生数	工科大、県立大、県立大夜間	3
学生男女比	工科大、県立大、県立大夜間	3
当月天気概要	降水量、平均気温、日照時間の例年 比それぞれ高中低で 2 値	9
前月天気概要	前月の降水量、平均気温、日照時間の例年比それぞれ高中低で 2 値	9
客数	その月の客数、午前午後毎	2
ライバル商品数	食べ物ならその月の同じジャンルの 商品数、飲み物ならその月の同じ味の 商品数	1
新商品	2 値	1
前月売り	その商品の前月売り有無	1
前月売上 個数	その商品の前月売上個数	1
合計		109

(表 2 商品売上予測特徴量)

## 5. 結果

結果については、特徴量・ハイパーパラメータの調整を複数回行った結果一番精度が良いと判断したものの詳細を記載す

る。

またこれから記載する結果は、特徴量及びハイパーパラメータを固定した上で5回、トレーニングデータ及び訓練データを変えて行った分析を行い、その中でMAEが中心になったものを用いた。結果の解釈であるが、MAEが小さいほど、予測と実測の散布図の近似直線が45度に近いほど、散布図の点が近似直線に近いほど実測値と予測値の差が小さく良いモデルといえる。

## 5-1 来店客数予測

### 5-1-1 精度

#### 5-1-1-1 初期結果

初期状態の学習・検証で得られた結果を以下に記載する

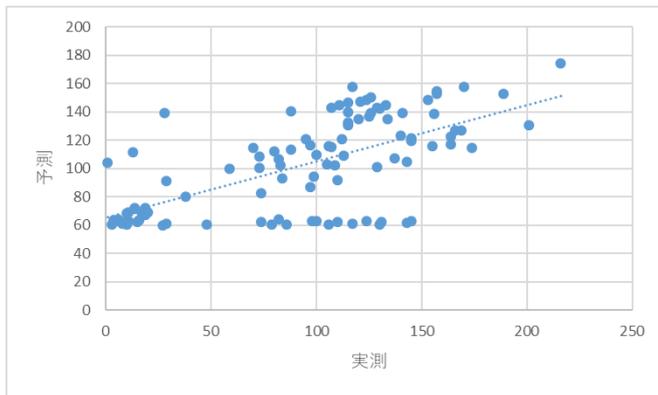
MAE = 32.955

MAEが32であるため、予測値と実測値では平均で32のずれが生じていることになる。

ハイパーパラメータ

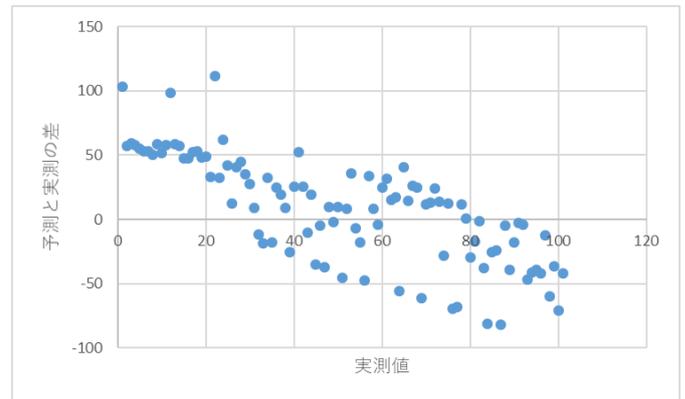
エポック数	ノード数	層
5	1024	2

予測と実測の散布図、点線は近似直線である。



(図5 来店客数 初期結果 予測散布図)

実測値と予測と実測の差の散布図である。



(図6 来店客数 初期結果 差散布図)

実測値が小さいほど大きく、実測値が大きいほど小さく予測している。実測値が平均から離れると精度が落ちている。

#### 5-1-1-2 最終結果

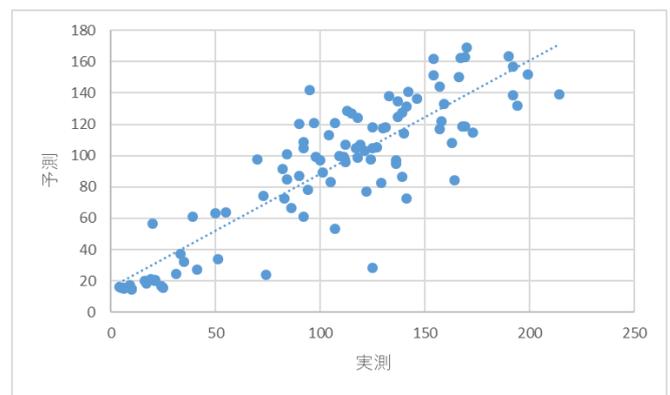
特徴量の調整・ハイパーパラメータの調整を行った後の学習・検証で得られた結果を以下に記載する。

MAE = 21.749

ハイパーパラメータ

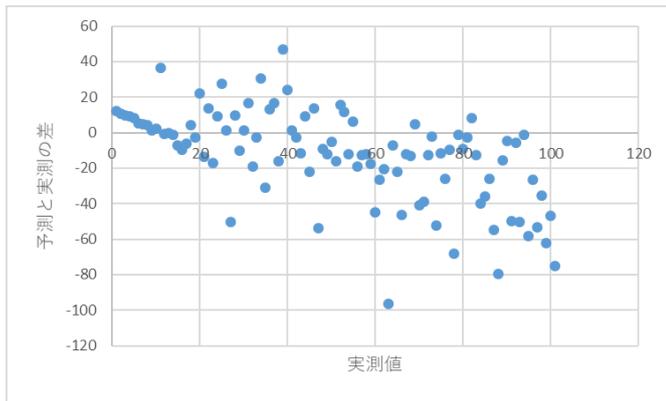
エポック数	ノード数	層
7	5000	1

予測と実測の散布図、点線は近似直線である。



(図7 来店客数 最終結果 予測散布図)

実測値と予測と実測の差の散布図である。



(図8 来店客数 最終結果 差散布図)

実測値が大きくなるほど予測の精度が落ちている。  
基本的に少なく予測する傾向がある。

### 5-1-2 特徴量調整

初期特徴量から調整を行ったもののみを記載する。

特徴量	変更・詳細
曜日	多重共線性排除のため「金曜日」削除 おなじく「土曜日」も排除
月	「12月」の項目を多重共線性排除のため削除
年	「2014年」の項目を多重共線性排除のため削除
前日 来店客	前日来店客数ではなく、前営業日の来店客数に変更。 前営業日午前・午後に高い相関が見られたため、二つを足し、「前営業日来店客数」を作成。
前営業日 補足	前営業日の降水量、授業の有無、テストの有無を追加
学生数	工科大学学生数と県立大学生に強い相関が見られたため、県立大学生数を削除 夜間県立大学生数を削除し、県立大に占める夜間学生の割合を追加
学生 男女比	県立大学生男女比と、県立大夜間学生男女比の間に強い相関が見られたため、県立大夜間学生男女比を削除

(表3 来店客数 最終特徴量)

### 5-1-3 重み

重みをみることで、AIがどの特徴量に重きを置いているのか

を判断することができる。

最終結果の重みを大きい順に5つ記載する。

1. 前日来店客数
2. 8月
3. 工科大授業
4. 11月
5. 2019年

特に重みが大きかったのは「前日来店客数」のみとなった。

### 5-1-4 考察

実測値が少ない時ほど予測の精度が高くなっている。  
実測値が大きくなるほど予測の精度が悪く、また少なく予測してしまう。来店客数が多く来る際の特徴量をモデルに含み切れなかったのが原因と考えられる。前日の来店客数ではなく、前営業日の来店客数に強く影響を受けることがわかった。  
(前営業日に関連する情報を入れると精度があがった)

## 5-2 商品売上予測

### 5-2-1 精度

#### 5-2-1-1 初期結果

初期状態の学習・検証で得られた結果を以下に記載する

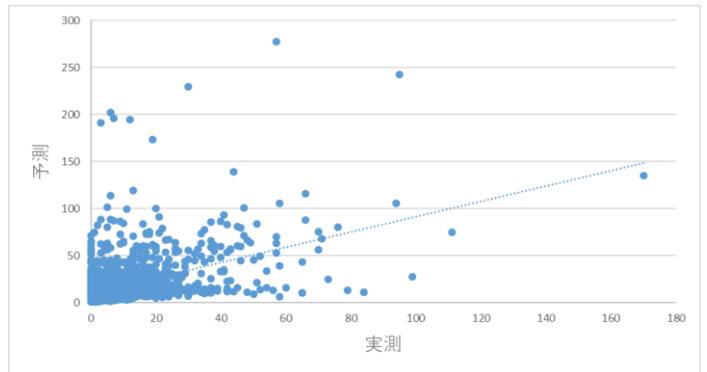
MAE = 11.028

誤差3未満割合 : 30%

ハイパーパラメータ

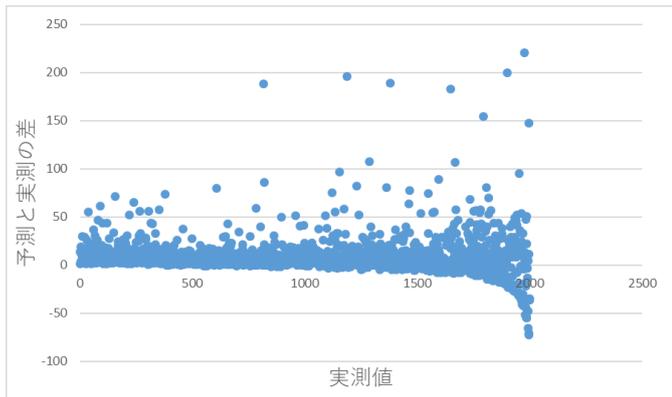
エポック数	ノード数	層
5	1024	2

予測と実測の散布図、点線は近似直線である。



(図9 商品売上予測 初期結果 予測散布図)

実測値と予測と実測の差の散布図である。



(図 10 商品売上予測 初期結果差散布図)

基本的に多めに予測する傾向がある。

### 5-2-1-2 最終結果

特徴量の調整・ハイパーパラメータの調整を行った後の学習・検証で得られた結果を以下に記載する。

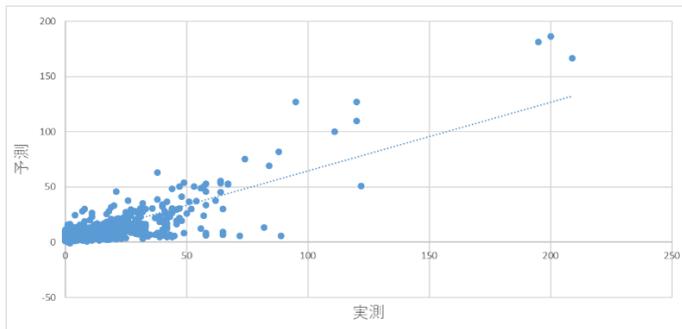
MAE = 4.857

誤差 3 未満割合 : 45%

ハイパーパラメータ

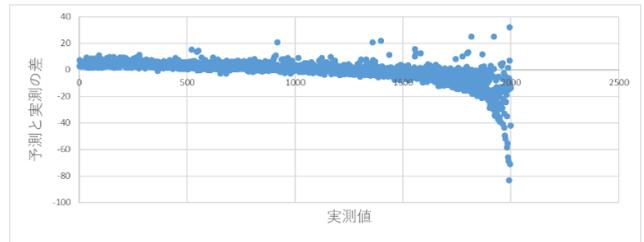
エポック数	ノード数	層
7	1024	2

予測と実測の散布図、点線は近似直線である。



(図 11 商品売上予測 最終結果 予測散布図)

実測値と予測と実測の差の散布図である。



(図 12 商品売上予測 最終結果 差散布図)

基本的には多めだが、実測値が大きくなると、少なめに予測してしまう。

### 5-2-2 特徴量調整

初期特徴量から調整を行ったもののみを記載する。

特徴量	変更・詳細
当月 天気概要	降水量、平均気温、日照時間の高、中、低をそれぞれ 2 値で表していたが、「高」を 2、「中」を 1、「低」を 0 とした連続値で表した方が、精度が向上したため、連続値に変更
前月 天気概要	上記、当月天気概要と同じく変更
月	「12 月」の項目を多重共線性排除のため削除
年	「2019 年」の項目を多重共線性排除のため削除
商品 ジャンル	56 個ある特徴量のうち 12 個を多重共線性排除のため削除
授業概要	工科大の授業数と県立大の授業数に強い相関が見られたため、工科大授業数を削除
学生数	工科大学学生数と県立大学生に強い相関が見られたため、県立大学生数を削除 夜間県立大学生数を削除し、県立大に占める夜間学生の割合を追加
学生 男女比	県立大学生男女比と、県立大夜間学生男女比の間に強い相関が見られたため、県立大夜間学生男女比を削除
客数	午前と午後に来店客数を分けていたが、強い相関が見られたので、二つを足して一日の来店客数に変更
新商品	MAE に正の影響を与えていると見られたため削除

(表 4 商品売上予測 最終特徴量)

### 5-2-3 重み

重みをみることで、AI がどの特徴量に重きを置いているのかを判断することができる。

最終結果の重みを大きい順に 5 つ記載する。

1. 前月売上
2. ライバル商品数
3. 工科大生徒数
4. 値段
5. 商品カテゴリー、スナック菓子

特に重みが大きかったのは、「前月売上」「ライバル商品数」の二つとなった。

### 5-2-4 考察

全体についてだが、実測値が大きくなるほど、予測値との開きが大きくなることから、予測する際に「売れる商品」及び「来店客数が多い日」に関連する要素を特徴量に含め切れていなかったことが考えられる。

### 5-3 実店舗での利用可能性

得られた結果を元に、永国寺ショップの担当者の方にお話を伺ったところ、「参考程度にはなるが、誤差・外れ値が大きく頼るのは不安」という回答が得られた。使える条件として提示されたのは、来店客数予測 MAE15 未満、売上予測は MAE3 未満割合が 70%以上という回答が得られ、実店舗での利用では精度の向上が必要であると感じた。また平均から大きく外れた値の予測精度が著しく悪いところも実店舗で利用するさいの留意点になる。

## 6. 考察と今後の課題

来店客数予測にしても、月毎売上予測にしても前日・前月のデータが大きく関係していることがわかった。平均から外れた値の予測の精度はわるくなってしまふ。外れ値の排除若しくは外れ値を起こしうる特徴量を捉えることができれば、予測の精度は上がる可能性がある。

生徒数や授業があるかどうかを AI が重要視していたので、生協ショップ独自のマーケティング属性が存在することが明らかになった。

今後の課題として来店客数予測は前営業日の値を特に重要

視しているため、前営業日の実態を正しく特徴量に組み込むことで精度の向上を図ることができると思う。

商品売上予測は、データが 1 か月毎のデータしか取得できず、さらにその商品が 1 か月間棚に置いてあり、その個数売れたのか、途中で在庫切れ若しくは棚から降ろしたのかが不明であったため、データの質が落ちてしまったのも予測を悪くした原因であると思う。よって、日毎の売上データを用い、予測する事が出来れば予測精度を向上させられるかもしれない。

## 7. 謝辞

本論文を作成するにあたり、ご指導を頂いた担当教員の那須清吾先生に心から感謝致します。また、日頃から多くの学びを授けてくださった先生方に深く感謝するとともに、ご意見やご提言を下された同期の皆様にもお礼申し上げます。またインタビュー及びデータのご提供を頂きました、高知工科大学永国寺キャンパス生協ショップのご担当者様にもお礼申し上げます。

## 8. 参考文献

・主要 121 社調査：AI 導入企業 47% 効率化へ研究進む - 毎日新聞

<https://mainichi.jp/articles/20180106/ddm/001/020/145000c>

(最終閲覧日 2020 年 1 月 28 日)

・機械学習と人工知能の違い | TechCrowd

<https://www.techcrowd.jp/machinelearning/ai/>

(最終閲覧日 2020 年 2 月 6 日)

・多重共線性と VIF 統計量の求め方 |  $\beta$  short Lab

<https://betashort-lab.com/データサイエンス/統計学/vif/>

(最終閲覧日 2020 年 2 月 10 日)

・精度評価指標と回帰モデルの評価 | Instruction of chemoinformatics

<https://funatsu-lab.github.io/open-course-ware/basic-theory/accuracy-index/>

(最終閲覧日 2020 年 2 月 10 日)