

共同研究「経営実務データを用いたデータサイエンス
テスト育成方法の研究」報告書

2021年3月31日

国立大学法人滋賀大学データサイエンス教育研究センター長

竹村 彰通



祝 辞

2021年3月に滋賀大学でデータサイエンス学部の1期生がご卒業されましたことに祝意を表します。

2017年にデータサイエンス学部が創設されてから、4年間共同研究を継続させて頂きました。本研究結果が微力ながら貴学部のデータサイエンティスト育成に資することができたことを喜ばしく思います。

今後とも、共同研究を通じて、データサイエンティスト育成を支援させて頂くとともに、卒業生の将来のご活躍を期待しております。

特定非営利活動法人 ビュー・コミュニケーションズ

1. 共同研究の趣旨

本共同研究の研究題目は「経営実務データを用いたデータサイエンティスト育成方法の研究」であり、本共同研究の目的は企業の経営実務データを用いて販売予測や適性な在庫管理など、これからのデータサイエンティストに求められる能力を育成するための方法を継続的に研究することである。本共同研究の期間は、令和2年11月2日から令和3年3月31日までである。

特定非営利活動法人ビュー・コミュニケーションズは実際の販売データを教育目的のために適切に処理をおこなった上で国立大学法人滋賀大学に提供し、滋賀大学データサイエンス教育研究センターはデータを2年生対象の「時系列解析入門」の講義の演習の形に整備して学生の教育に活用することで、データサイエンティスト育成方法に関する共同研究をおこなった。

本年度は昨年度に引き続きビュー・コミュニケーションズから、食品スーパーにおける5年分の酒類の月次販売データが提供された。このデータは滋賀大学のデータサイエンス学部の学生にとって、ビッグデータの一端を実感できる実務データであり、時系列解析入門の講義でのレポート課題として用いることにより、学生達は講義で学習する時系列解析の手法がどのように実際のデータ分析に役立つかについて理解を深めることができた。このような実際のデータを用いた演習は本共同研究の目的に沿うものであり、滋賀大学データサイエンス学部が掲げる実践的なデータサイエンティスト育成の観点からも非常に有用であった。その点で本共同研究の意義は大きい。

2. 滋賀大学データサイエンス学部の育成人材像と教育実績

滋賀大学データサイエンス学部では、データサイエンスの基礎要素技術としてのデータエンジニアリング(情報工学)及びデータアナリシス(統計学)に加えて、データからの価値創造の能力の育成を重視している。価値創造の能力を育成するためには、実際のデータを用いた演習の中で、試行錯誤や成功体験が必要であり、さらに教科書的な理論と実際のデータとの乖離についても経験を積む必要がある。理論と実際の乖離については、ビュー・コミュニケーションズ副理事長の小松秀樹氏の著書「なぜあなたの予測は外れるのかーAIが起こすデータサイエンス革命」においても多くの例とともに示されており、データサイエンス教育において実際のデータを扱うことが非常に重要であることがわかる。

滋賀大学データサイエンス学部は日本初のデータサイエンス学部として注目をあびていたが、2017年4月の開設以来この3月にはいよいよ1期生が卒業となった。卒業生の就職状況はかなり良く、データサイエンス分野に関する社会のニーズが強いことを示している。また、価値創造に重点をおいてきた学部の教育方針は、例えば卒業論文のテーマにもよく現れている。

2021年2月4日、5日の二日間に渡って、データサイエンス学部第1期生の学習の総括

として、卒業論文の発表会を実施した。発表会では、事前に各学生が提出した卒業論文の内容を各自が口頭発表し、質疑を行った。発表会はオンライン会議システム ZOOM で発表資料を画面共有することで実施し、発表時間は 1 人につき、発表 7 分質疑 3 分の計 10 分とした。卒業論文の内容が、グループで研究の一部のこともあったが、この発表会では各自の貢献箇所を中心に学生が個別に発表するものとした。今回の発表会では、93 名の発表者を 11 のセッションに分け、そのうちの 1 つのセッションは、指導教員から推薦された発表を集めた特別セッションとした。

特別セッションの発表者(指導教員)と発表タイトルの一覧は表の通りである。特別セッションでは、学部連携企業の方々を外部評価委員としてお招きし、質疑に加わって頂いた。評価委員からは、「実際のデータから実際の施策立案まで落とし込めたことは非常に良かったのではないかと思います」「内容、論理性、一貫性はとても高い水準だったと思います」「身近な話題を抑場のある巧みな説明で上手に伝えられていた」などのコメントがあり、活発に質疑が行われた。

発表者 (指導教員)	発表タイトル
森本 濤二 (佐藤 智和)	実世界の仮想化に基づく高臨場 VR 型防災教育システムの開発
水口 綾乃 (市川 治)	事前学習済み分散表現を利用した学部オープンキャンパス向け質問 応答システムの構築
江口 公基 (加藤 博和)	彦根市を目的地とした観光交通における鉄道利用促進のための機関 選択分析
高田 拓弥 (松井 秀俊)	関数データに基づく回帰モデルと農業・化学分野への応用
森口 翼 (河本 薫)	テナント型商業施設における会員用スマホアプリのログデータ分析 による離反防止策の検討
田室 建志 (河本 薫)	自動車部品工場における機械学習を活用した異常検知モデルの構築
上田 知展 (清水 昌平)	平和堂の ID 付き POS データを活用したモバイルクーポンの改善
小西 秀明 (清水 昌平)	ID 付き POS データを利用したモバイルクーポンの仕様改善に関する 施策提案と効果検証
谷口 友哉 (清水 昌平)	購買履歴データを用いたモバイルクーポン配信の最適化
仲北 昌大 (和泉 志津恵)	2018 年 7 月の西日本豪雨災害のアンケートデータから分かる発見と 問題点

この特別セッションの発表題目にみられるように、多くの卒業論文の内容は、企業等の現実のデータを分析しそこから価値を引き出すものであった。外部評価委員のコメントにみられるように、これらの内容は企業の視点から見ても実践的なものであった。

このように1期生の卒業論文には滋賀大学データサイエンス学部の教育の成果が反映されているが、このような教育方針を今後も深化させていく必要がある。ビュー・コミュニケーションズから提供されたデータを用いた演習教材のような、価値創造につながる教育コンテンツの開発は滋賀大学が今後も展開していくべき活動である。

3. 滋賀大学大学院データサイエンス研究科の教育実績と拠点形成

滋賀大学ではデータサイエンス教育を大学の戦略的な方針として位置付けており、データサイエンス教育研究拠点の整備を進めてきた。2019年4月には、日本初のデータサイエンス研究科修士課程を開設し、この3月には学部の1期生と同時に修士課程の1期生も卒業となった。修士課程の開設は、データサイエンス学部の開設から2年後のことであり、まだ学部からの卒業生が出ていない段階での「前倒し設置」であり、学部からの進学性がいないため入学者は外部からの応募者であった。定員20名のところを23名が修士課程1期生として入学したが、そのうち19名は企業派遣の社会人であった。これらの社会人は、それぞれ派遣元の企業の実際の課題を解決するための修士論文研究をおこない、2021年2月10日及び12日の二日間に渡って修士論文発表をおこなった。これらの研究の多くは、直接派遣元の企業の業績向上につながるものである。

また2020年4月には博士後期課程も開設した。博士後期課程は当初の定員は3名と少数だが、いずれも企業所属の3名が入学し、博士課程での研究を通じて棟梁レベルのデータサイエンティストとなることを目指している。このように滋賀大学では学部から博士課程まで一貫したデータサイエンス教育体制が完成した。

4. 本年度ビュー・コミュニケーションズより提供されたデータ

本年度にビュー・コミュニケーションズより提供されたデータは、昨年度と同様に関東圏食品スーパー100店舗サンプルの販売金額合計値時系列データであった。販売品目は酒類10品目であった。データはエクセルのファイルとして、すでに整形された形で提供されており、学生はデータをRやPythonなどの統計分析パッケージに読み込むことで、ARIMAモデルなどの時系列解析の手法を応用することができた。このデータは時系列解析の例題としては非常に実践的なものであった。次図は提供されたエクセルファイルの一部を示したものである。

	第3のビール	ビール 350ml	焼酎甲類	清酒パック	国産ワイン	輸入ワイン	国産ウイスキー
2014年10月	64,896,265	56,458,480	41,283,034	27,710,083	15,671,871	17,012,474	11,630,853
2014年11月	62,626,576	53,164,431	42,021,722	30,565,834	17,547,875	33,758,505	12,350,601
2014年12月	61,599,319	73,810,979	44,625,092	37,739,137	20,952,974	25,209,042	14,292,028
2015年1月	56,903,102	54,636,556	39,416,436	28,568,680	15,394,759	18,412,583	13,251,262
2015年2月	55,275,135	42,429,178	37,706,181	28,192,179	14,140,309	16,260,936	13,442,114
2015年3月	61,832,614	55,360,190	41,023,819	28,121,201	15,395,572	16,849,718	15,683,896
2015年4月	65,137,971	55,137,113	39,079,248	25,829,663	13,980,098	16,657,483	15,533,525
2015年5月	95,580,312	75,617,063	43,269,560	22,720,433	13,507,734	17,591,824	14,537,782
2015年6月	88,410,801	64,505,572	55,492,424	20,548,710	11,904,098	15,005,699	13,854,597
2015年7月	96,231,198	77,867,879	39,422,324	19,173,130	11,388,621	14,482,564	13,158,490
2015年8月	95,740,071	93,823,473	38,529,894	18,322,581	10,929,671	14,653,894	13,122,023
2015年9月	85,365,554	66,422,417	38,211,951	22,253,526	12,521,920	15,458,816	12,324,527
2015年10月	87,278,610	61,513,467	40,727,491	27,928,000	16,921,779	19,699,321	13,375,270
2015年11月	80,622,982	57,462,902	41,008,635	31,193,441	17,679,276	34,439,141	13,335,337
2015年12月	80,550,355	84,451,103	43,382,753	37,712,380	20,849,978	26,678,527	15,593,571
2016年1月	74,896,265	63,204,148	39,193,684	30,669,160	15,611,565	19,267,511	13,578,948
2016年2月	76,489,294	49,531,205	37,564,149	29,950,650	14,643,939	18,477,289	13,249,479
2016年3月	84,751,597	64,607,404	40,332,548	30,191,077	14,764,217	19,224,918	14,284,018
2016年4月	85,530,588	66,563,713	38,443,524	27,844,328	14,134,252	19,286,489	15,331,135
2016年5月	94,723,466	78,050,551	42,703,326	25,602,536	14,566,449	21,522,904	16,997,167
2016年6月	90,770,895	71,194,568	51,544,702	22,761,122	12,312,624	17,275,704	18,719,842
2016年7月	99,618,096	84,006,041	39,677,949	21,045,965	11,559,434	16,372,270	19,395,526
2016年8月	97,794,657	99,688,086	38,734,847	19,953,496	10,942,939	15,548,213	19,338,026
2016年9月	89,843,497	67,595,129	37,984,195	22,357,293	12,079,329	16,311,333	19,512,245

本年度提供データの一部

5. 学生によるデータ分析の例

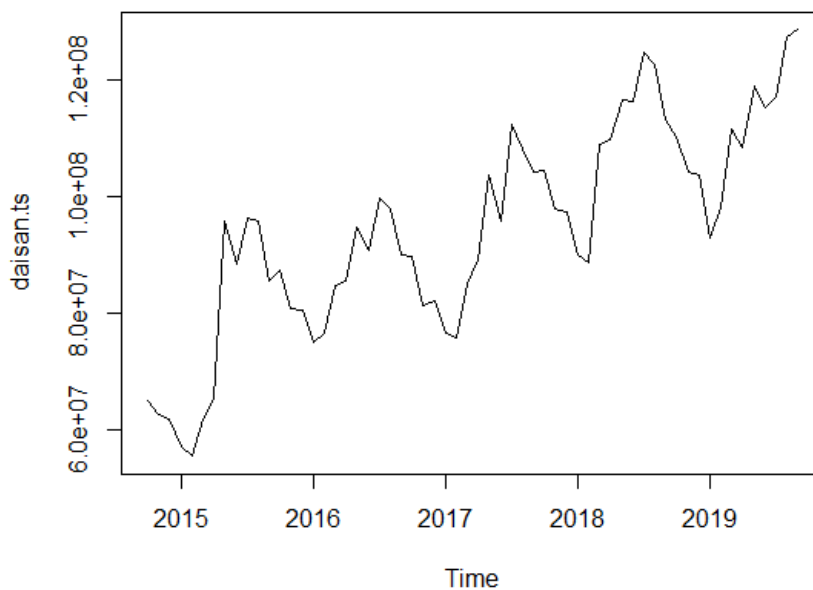
レポート課題では、ビュー・コミュニケーションズより提供された10品目のデータの中から3品目を選んで解析するよう学生に求めた。以下は第3のビールに関する浅野菜月さんの解析例である。

(1) 第3のビールについて

```
daisan <- scan("daisan.dat")
```

```
daisan.ts <- ts(daisan, st = c(2014,10), end = c(2019,9), fr = 12)
```

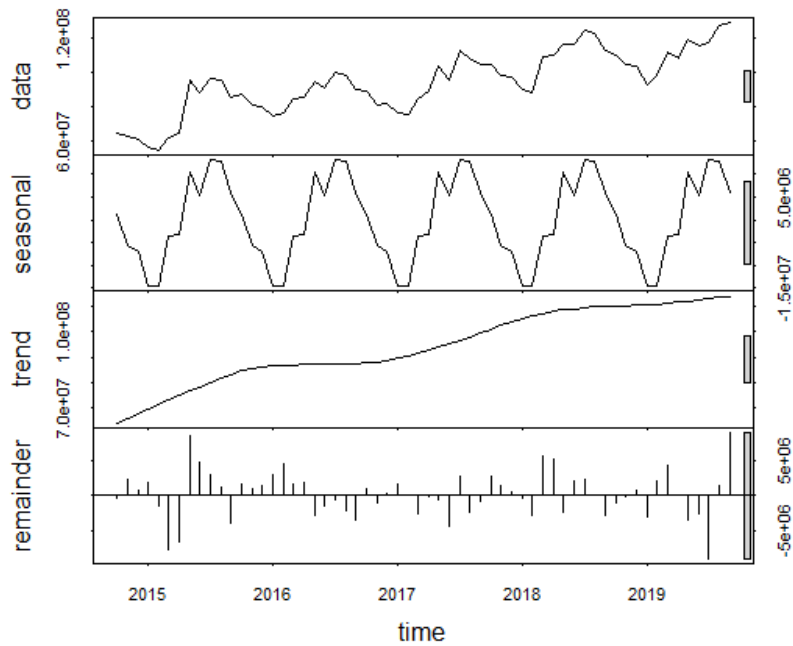
```
plot(daisan.ts)
```



```
daisan.stl<-stl(daisan.ts,s.window="per")
```

```
plot(daisan.stl)
```

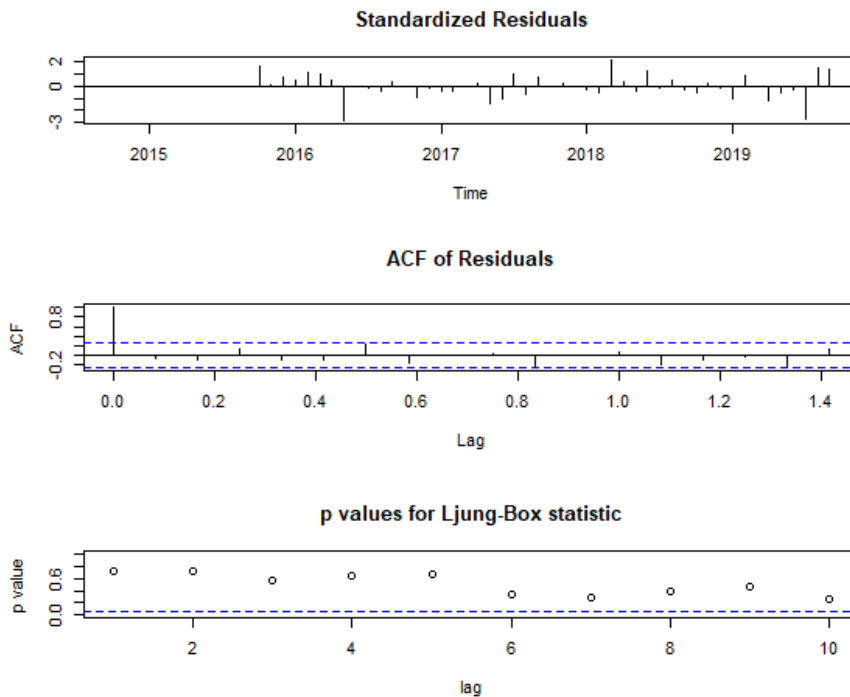
推定された傾向成分より、第3のピールは近年増加傾向にあることが分かる。



```
daisan.arima<-auto.arima(daisan.ts,stepwise=T,trace=T)
```

```
tsdiag(daisan.arima)
```

以下の結果から、自己相関も小さく、誤差らしい誤差（ホワイトノイズ）であるということが分かり、よいモデルといえる。



1 年先の予測をしてみると、

```
daisan.pred<-forecast(daisan.arima,level = 95,h=12)
```

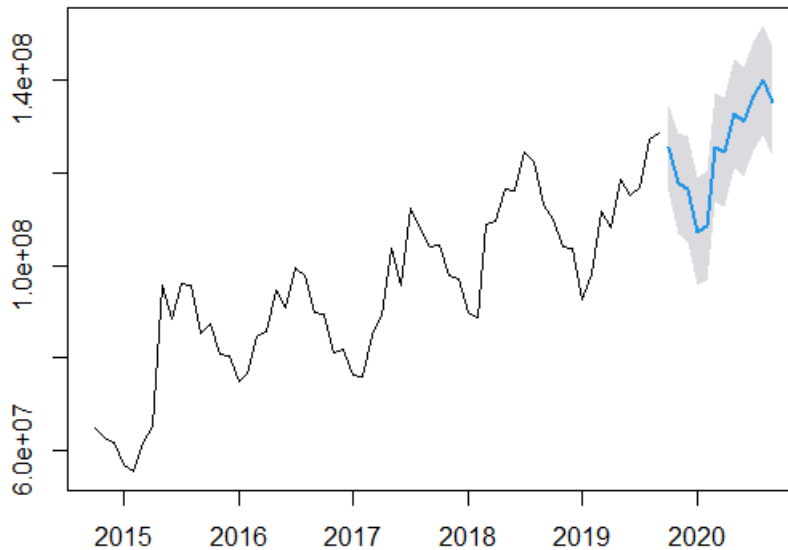
```
class(daisan.pred)
```

```
daisan.pred
```

	Point Forecast	Lo 95	Hi 95
Oct 2019	125419363	116068454	134770271
Nov 2019	117881430	106918635	128844226
Dec 2019	116599663	105091219	128108106
Jan 2020	107263155	95556927	118969384
Feb 2020	108559279	96779841	120338717
Mar 2020	125676803	113870067	137483539
Apr 2020	124585109	112768167	136402051
May 2020	132980797	121160035	144801559
Jun 2020	131187141	119364950	143009333
Jul 2020	136690302	124867575	148513029
Aug 2020	140030733	128207806	151853661
Sep 2020	135191841	123368838	147014844

```
plot(daisan.pred)
```


Forecasts from ARIMA(1,0,0)(1,1,0)[12] with drift



1年先もこれまでと同じような季節の影響等を受けながら、増加していくということが分かる。グレーの幅が信頼区間を示しているが、このグラフでは全体的に幅がほぼ同じように見えるが、下に大きく下がっている部分など所々幅が大きいところも見られる為、大きく増減しているところに関しては少し推定の精度が低いといえるのではないだろうか。

このように、浅野さんの分析は、データのプロットから始め、Rのforecastパッケージで提供されているauto.arimaコマンドを用いて最適なSARIMAモデルを選び、選ばれたモデルの当てはまりまでを検討しており、標準的な分析例となっている。

6. まとめ

前節でみたように、提供されたデータの分析のレポート課題に多くの学生が熱心に取り組んだ。その理由は、提供されたデータが非常に実践的で、また時系列解析の効果が実感できるものであったことである。

滋賀大学データサイエンス学部としては、今後もさらに興味深い演習課題をビュー・コミュニケーションズとの共同研究の形で開発して行きたいと考えている。修士課程の教育にも活用していきたい。さらに、滋賀大学は数理及びデータサイエンスに係る教育強化の拠点校として文部科学省より選定されており、このような教育コンテンツを他大学に展開していくことも重要な課題である。