

# 農産物向け生産データ管理システムの開発とデータの活用

工藤 智子\*・庄司 光平\*・五十嵐 萌衣\*・金内 諒\*・本間 祐哉\*・成田 浩輝\*\*

## Development of a data management system for agricultural products and utilization of the data

Tomoko Kudo\*, Kohei Shoji\*, Moe Igarashi\*, Ryo Kanauchi\*, Yuya Honma\* and Hiroki Narita\*\*

**要旨：** 筆者らは共同研究として、庄内地方でのベビーリーフ生産のための農産物生産データ管理システムの開発に取り組んできた。また、システムに蓄積したデータを活用し、機械学習を活用した収穫量予測のための分析を行った。本稿では、2018年度からの共同研究の活動報告を行う。

**キーワード：** 農業, Society5.0, 生産量予測, 重回帰分析, 機械学習

### 1. 緒言

農業分野における Society5.0 の実現として、スマート農業がすすめられている。その中で、情報共有の簡素化（作業の記録のデジタル化）や自動化、データの活用をすることにより高度な農業経営が可能になるといわれている 1)。しかし 2)によれば、山形県内の農業経営体数 5,769 体のうち、気象や生育履歴等のデータを活用した農業経営を行っている割合は 20.4%、さらにそれらを分析して活用する経営体は 1.2%（団体経営体で 6.4%）にとどまっており、データ活用のすそ野を広げることが急務と考えられる。

庄内地方で在来作物を生かしたベビーリーフの栽培と販売を行っている株式会社いで葉工望は、在来作物を生かしたベビーリーフの栽培と販売を行っている。筆者らは、生産データ管理システムの開発手法やシステムに蓄積されるデータを活用した分析手法について検討を行ってきた。

庄内地方は、冬期間の日照量が県内でも少なく、特に今年度の日照不足は野菜類の生育に大きな影響を与えている。その一方、消費者からの県内産野菜の需要は大きく、特に冬季需要期の生育状況の把握は喫緊の課題となっている。

ベビーリーフの大きな特徴として、多品種を少量ずつ生産し、パッケージ化して出荷するが、多品種少量生産農産物に適したサービスやパッケージソフトの導入には、カスタマイズやノウハウが必要である。

将来的に品種毎の生産計画を立てられることを目標とし、そのために必要となるデータの蓄積や集計作業の効率化を図るため、「ベビーリーフ生産データ管理システム」の設計から開発を行った。

さらに、生産データ管理システムに蓄積したデータを活用し、冬季需要期におけるベビーリーフの生育の予測に生かすことができるか、機械学習による収穫量の予測を試みた。本報告では、これまで取り組んできたシステム開発とデータを活用した分析についての報告を行う。

### 2. ベビーリーフについて

ベビーリーフ（図 1）とは、発芽後 10～60 日程度で収穫した野菜やハーブの若葉のことである。ベビーリーフは、ミズナ、ピノグリーン他約 20 種を少量ずつ生産し、複数の品種から数種類を選択して、決められた重さにパッケージ化し出荷する（図 2）。



図 1 ベビーリーフ

ベビーリーフは、暖房等で温度制御をしないハウ

\*山形県立産業技術短期大学校庄内校  
〒998-0102 山形県酒田市京田三丁目 57-4  
e-mail: tomoko@shonai-cit.ac.jp

\* Shonai College of Industry & Technology  
3-57-4 Kyoden, Sakata City, Yamagata, 998-0102, Japan  
e-mail: [tomoko@shonai-cit.ac.jp](mailto:tomoko@shonai-cit.ac.jp)

\*\*株式会社いで葉工望  
〒999-7781 山形県東田川郡庄内町余目上朝丸 74-1  
74-1 Kamiasamaru, Amarume, Shonai Town, Yamagata,  
999-7781, Japan

ス内で栽培しており（図 3）、比較的安定した環境ではあるものの、特に日照量が減少する冬期間は、播種から刈取までの期間が長くなる。特に年末の需要期の収量確保が課題であり、生産・選別・出荷に関するデータを活用し、生産計画をたてる必要がある。



図 2 播種から出荷工程



図 3 ベビーリーフ栽培用ハウス(庄内町)

### 3. システム開発手法の検討

#### 3.1 要件定義～データベース設計～インフラ構築 (2018 年度)

共同研究開始の初年度である 2018 年度は、生産データ管理システムの要求定義を行った。

- (1) 業務と業務フロー, 利用者, 利用場所
- (2) システムに求められる機能
- (3) システム化の範囲
- (4) 制約条件
- (5) 運用方法

について定義し、システムの外部設計から開発を行った。また、ネットワークインフラとして、クラウドサービスである Amazon Web Services™ を利用することにより、事務所や、将来的には圃場からの利用ができ、バックアップ等のセキュリティ対策が行えるようにした。

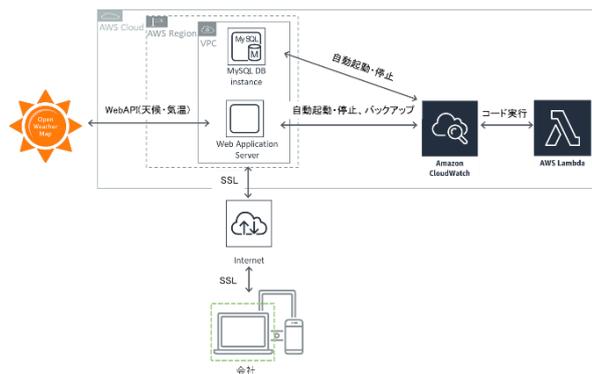


図 4 クラウドサービスを利用したシステム構成 (2018 年度)

表 1 利用したクラウドサービス

Cloud Service	Amazon Web Services™
Web Application Server	Amazon EC2™
Database Server	Amazon RDS™
Computing Service	AWS Lambda™
Monitoring Service	Amazon CloudWatch™

#### 3.2 開発手法の検討 (2019 年度)

2018 年の開発の際には、クラウド上に Web アプリケーションを実装したが、今後のシステム拡張の可能性を考慮し、「ローコード開発」利用の検討を行った。ローコード開発とは、コードをほぼ書かずにアプリケーションを開発する手法で、高速にアプリケーションの開発が可能であるといわれる手法である。複数のローコード開発ツールがある中、OutSystems™ を利用し、開発時間の削減の効果があつたことを確認した。

開発はデータ駆動により行うため、最初にデータベースのテーブルとそのリレーションを設定する（図 5）。次に、必要な部品やデータベーステーブルを挿入することにより、必要な画面を自動的に作成する（図 6）。レスポンス Web デザインのサイトが作成されるため、デスクトップ用とスマートフォン用アプリが同時に作成可能である（図 7）。その一方、複数テーブルから画面を生成する場合には、ノウハウが必要である。画面遷移が多くなる傾向があり、実際のシステム利用には UI の改善とカスタマイズが必要であつたため、今回は採用を見送ることにしたが、開発高速化には有効な手法である。

### 3.3 フロントエンド部分の構築（2020年度）

これまでシステム開発手法を検討してきたが、開発期間の短縮と運用負担の軽減を図るため、フロントエンド部分を Microsoft Access™ で再構成することにした。今後、バックエンドで動作している AWS との接続も可能な状況となっている。

本システムにより品種毎の刈取量・選別後収量、ロス率についてのデータの蓄積や見える化が実現している。

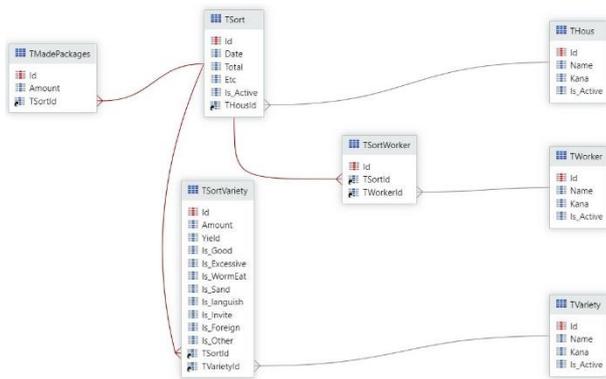


図 5 テーブルのリレーション設定

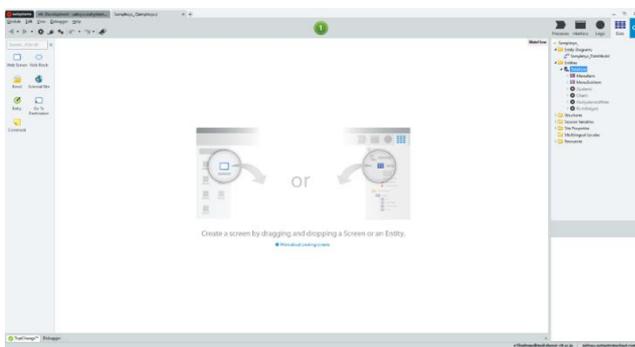


図 6 ローコード開発ツールによるシステム開発



図 7 ディスクトップ用(左)とスマホ用アプリ(右)



図 8 2020年開発システムの概要



図 9 2020年開発システム(データ活用画面)

### 4. データ分析と収穫予測に向けた分析手法の検討

庄内地方は、平均気温、平均日照時間ともに冬に低下する(図 10)。特に冬の日照時間は県内他都市と比較しても少なく、1日当たりの日照時間は1時間程度にとどまっている(図 11)。ベビーリーフはハウス内での栽培が行われているが、特に日照不足は野菜類の生育期間に大きく影響すると考えられ、冬季需要期の生育状況の把握が課題である。

そこで、気象データから刈取量、ロス率の予測が可能か検討する。表 2 に示す生産データと気象データから、過去3年間の品種毎の刈取量の推移やロス率の傾向を分析、さらに気象データとの相関を分析した。今後、特に年末年始のベビーリーフの冬季需要期に刈取量やロス率が予測可能か検討を行った。

まずデータの集計のために、表 2 のデータを開発

システムに入力した。これに、ビジネスインテリジェンス用ツール Tableau<sup>TM</sup>を用いて分析を行った。

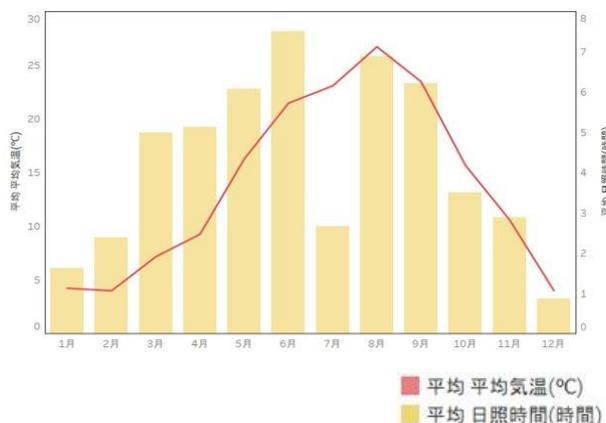


図 10 月別平均日照時間, 平均気温(2020年酒田市)

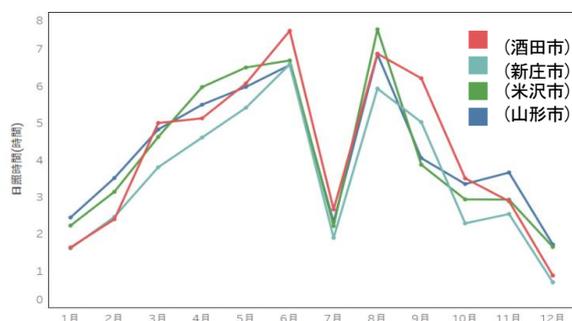


図 11 県内各都市の日照量平均(2020年)

表 2 分析対象データ

データ取得期間	2018年10月～2021年1月
分析対象期間	2018年10月～2019年1月 2019年12月～2021年1月
データ件数	4058件
分析対象データ	品種毎の刈取量 品種毎の選別ロス量
気象データ	日別の平均気温, 最高気温, 最低気温, 日照時間, 平均相 対湿度, 平均現地気圧 (気象庁 酒田市)

#### 4.1 気象データとの相関分析

気象データに対する刈取量・ロス率の相関の有無を求める。相関を分析するのは、(1)積算日照時間・平均気温が、ベビーリーフの発芽や成長への影響があるか、(2)積算気温や湿度等によりベビーリーフの品質への影響があるかを分析する。研究当初は、表 2 に示す全てのデータを用いて分析を行っていたが、有意な結果が得られなかったため、

2020年12月～1月のデータに限定し、分析を行った。その理由は、2020年から生産量が安定していること、6,7月のコロナ禍による刈取量の低下や廃棄増の影響を排除した。また、年末需要が高い時期の刈取量の予測を目標とするためである。また、主要6品種とは、年間を通じて刈取量の多いミズナ、ピノグリーン、グリーンマスタード、レッドマスタード、タアサイ、ルッコラである。

図 12 は、刈取量・ロス率と湿度・積算気温・日照時間との相関を示す散布図、図 13、図 14 は、刈取量・ロス率に対する積算気温の日数による相関の変化を比較したグラフである。

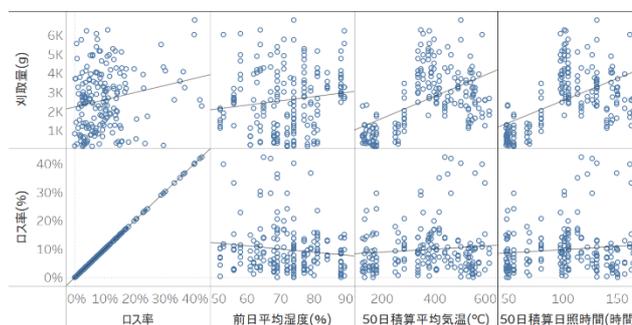


図 12 刈取量・ロス率と湿度, 積算気温・日照時間との相関(主要6品種 2020年12月～2021年1月)

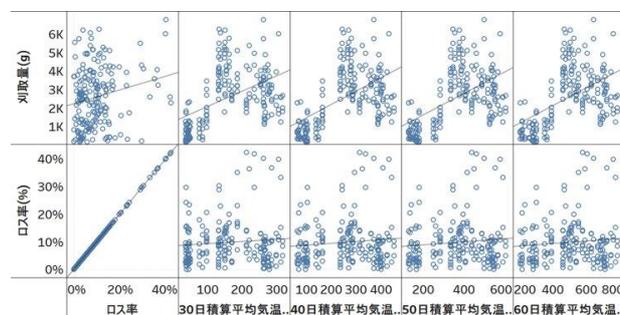


図 13 刈取量・ロス率と積算気温日数との相関(主要6品種 2020年12月～2021年1月)

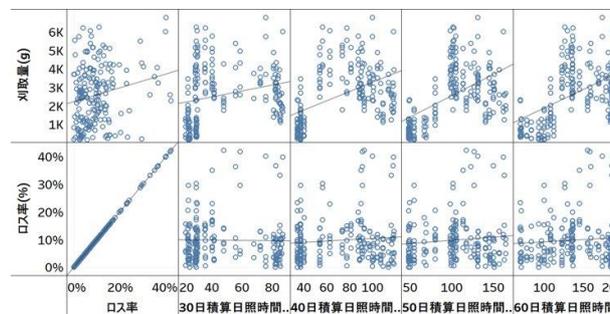


図 14 刈取量・ロス率と積算日照日数との相関(主要6品種 2020年12月～2021年1月)

これらの結果から、式 (1) に示す相関係数 (r

値) を求める。相関係数とは、2 種類のデータ  $x, y$  の関係性の強さを -1 から 1 の間で表した値である。

$$r = \frac{S_{xy}}{S_x \times S_y} \dots (1)$$

$S_{xy}$   $x, y$  の共分散,  $S_x, S_y$   $x, y$  の標準偏差

また, 式 (2) に示す決定係数 ( $r^2$  乗値) は  $x$  が  $y$  を決定する強弱の度合いを測る値である。

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots (2)$$

表 3 主要 6 品種における刈取量・ロス率と前日平均湿度, 50 日積算平均気温・日照時間の相関係数

	平均湿度 (前日)	積算平均気温 (50 日)	積算日照時間 (50 日)
刈取量	0.28	0.42	0.47
ロス率	0.05	0.33	0.34

表 3 は、主要 6 品種における刈取量とロス率と前日の平均湿度, 50 日積算平均気温, 日照時間に対する相関係数の比較である。刈取量に対して, 50 日積算平均気温は 0.42, 日照時間は 0.47 とやや強い正の相関がある。湿度による品質低下も考えられたが, 実際にはロス率と湿度に相関はなかった。この結果から, 正の相関がみられたのが積算平均気温・日照時間であったため, 図 15 に, 日照時間・平均気温について積算日数を変化させたときの刈取量に対する相関係数を比較したものを示す。積算日数 30 日から 50 日にかけて相関が高くなっており日照時間が増すと刈取量も増加する。全体的に弱い正の相関であり, 60 日になると若干相関が弱くなっている。また, 品種毎においても同様の比較を行っているが, 相関係数が強く出ている品種が存在する。

同様に, 図 16 に日照時間・平均気温について積算日数を変化させたときのロス率に対する相関係数の比較を示したである。全体的に弱い正の相関であるものの, 積算日数が増加してもほとんど変動は見られなかった。

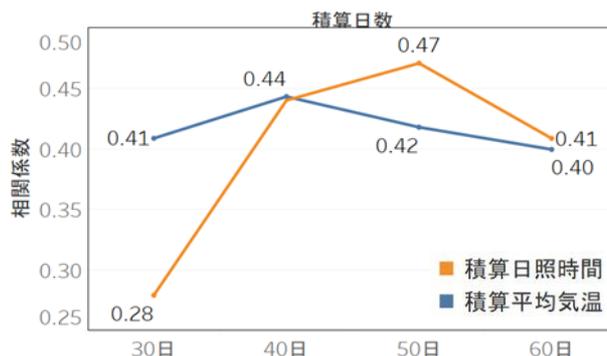


図 15 刈取量と積算日照時間・平均気温の相関係数 (主要 6 品種: 2020 年 12 月～2021 年 1 月)

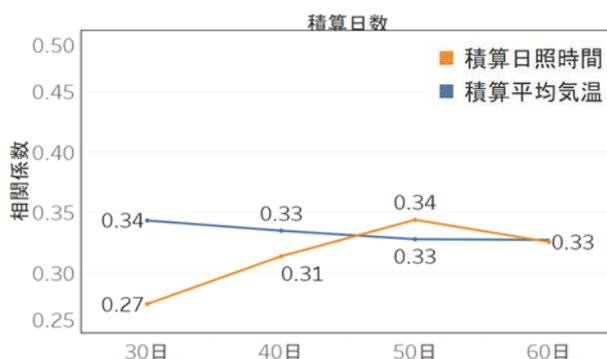


図 16 ロス率と積算日照時間・平均気温の相関係数 (主要 6 品種: 2020 年 12 月～2021 年 1 月)

#### 4.2 重回帰分析による生産データの予測

次に, 機械学習の一種である重回帰分析を用いて, 気象データから収穫量やロス率が予測するための学習モデルの作成とその検証を行った。

これは, 訓練データとして与える説明変数  $x_1 \sim x_n$  から目的変数  $y$  を予測するモデルを求める手法である。モデルに訓練データの説明変数を入力し, そのモデルからの出力が訓練データの目的変数に近づくようにモデルのパラメータ (係数や切片) を調整することで学習をすすめる。その結果, 各説明変数の係数 (回帰係数) が推定されて予測値  $y$  を計算できる (式 (3))。回帰係数は予測値と目的変数の二乗誤差が最小になるように推定される。データの一部を訓練データとして学習させ, 構築したモデルに残りのデータをモデル検証用のテストデータとして与え, その説明変数から目的変数の予測精度を確認する。予測精度の指標は 4.1 同様, 式 (2) に示す決定係数を使用して判断する。

$$y = a_1 x_1 + a_2 x_2 + \dots + a_n x_n + b \dots (3)$$

$x_1 \sim x_n$ : 説明変数  $a_1 \sim a_n$ : 偏回帰係数  $b$ : 切片

ここで, 予測する目的変数は, 刈取量とロス率, 説明変数は, 気象データ (平均気温, 最高気温, 最低気温, 前日平均湿度, 前日平均気圧, 日照時

間、各積算気温、各積算日照時間)を組み合わせて試行することとした。分析にあたっては、Pythonの実行環境である Google Colaboratory™を利用し、モデルの構築と評価を行った。

モデルの評価は、訓練データに対する決定係数とテストデータに対する決定係数が同じ傾向を示せばモデルの精度は高いと判断する。図 17 は、気温・日照時間の積算日数に対する決定係数を比較したものである。

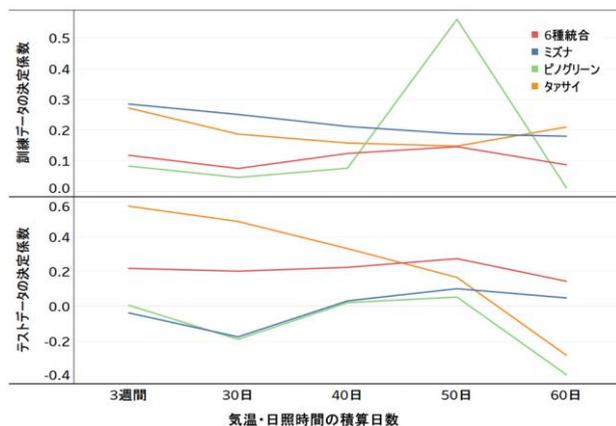


図 17 気温・日照時間の積算日数に対する刈取量の決定係数の比較(2020年12月~1月)

単品種のデータで作成したモデルは訓練データとテストデータの決定係数に乖離が見られる。テストデータの決定係数が低く、また、データ数の不足によるものと推測するが、学習不足の状態に陥ったと考えられる。6種混合(主要6品種の合計)は、単品種と比較すればまだ良いが、予測に用いるには十分とは言えない。また、積算日数に着目すると、ピノグリーンを除き、50日で訓練データとテストデータの決定係数が近くなっている。冬季の播種から刈取までの日数に近いため妥当な結果といえる。

また、図 18 は気温・日照時間の積算日数のロス率の決定係数の比較をしたものである。テストデータにおける決定係数がマイナスになる場合が多く、ロス率と気温・日照時間の積算に関連性は得られなかった。

さらに正確な予測モデルを作成するために、生産データ・環境データの蓄積、ハウス内環境の計測(室温・土壌温度・湿度)が必要である。

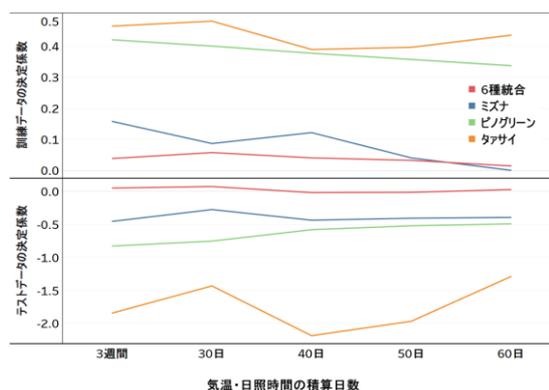


図 18 気温・日照時間の積算日数に対するロス率の決定係数の比較(2020年12月~1月)

## 5. 結言

本報告では、将来的に品種毎の生産計画を立てられることを目標とし、生産データ管理システムの開発とそのデータの分析を行った。

その結果、気温や日照時間が収穫量に対して相関が得られることは分かったが、精度の高い予測モデルの作成には至らなかった。今後、生産データの継続した蓄積と IoT を活用した各種データ取得が必須である。データ量が蓄積されれば、AI を活用した収穫量の予測、加えて、生産量の落ち込む冬季に備えた生産計画を立てるためのデータ活用が期待できる。

## 謝辞

共同研究をすすめるにあたり、卒業研究として取り組んでくれた山形県立産業技術短期大学校情報システム科卒業生青木浩亮さん、金子龍太郎さん、鈴木里実さん、榎谷勇介さん、山形県立産業技術短期大学校庄内校電子情報科卒業生伊藤颯さん、齋藤優也さんに心より感謝します。

## 文献

- 1) 農林水産省 スマート農業の展開について (2021)
- 2) 農林水産省 2020 年農林業センサス農林業経営体調査結果の概要(山形県分) (2021)