

# 機械学習サービスシステム開発プロジェクトにおける危険な匂いと実践に基づく収集

竹内 広宜 今崎 耕太 本橋 洋介 久野 倫義 土肥 拓生

機械学習 (Machine Learning: ML) を活用したサービスシステムの開発が近年盛んに行われている。一方、サービスシステムの開発ではプロジェクトを進める上で遭遇する技術的な課題を解決するため、様々な知見を蓄積し、別のプロジェクトの実施において活用している。ML サービスシステムの開発プロジェクト (ML プロジェクト) の実施においても、様々な知見が収集されており、ベストプラクティスやパターンとして整備されつつある。しかしながら、整備されたベストプラクティスやパターンはどういった場面で利用できるのかが不明確であるため、これらの体系化された知識の活用はプロジェクト担当者の経験やスキルに依存する状態にある。本研究では、既存のベストプラクティス (解決策) から、それを適用すべきプロジェクトの状態を危険な匂いとして収集することを検討した。そして、オンラインワークショップを通して実務家から危険な匂いを収集した結果を報告する。

In this study, we consider projects where we develop service systems using machine learning (ML) techniques. In various domains, ML techniques are introduced and as a result there is a reusable knowledge on the ML projects for conducting such projects effectively. In the reusable knowledge such as best practices or patterns, it is not clearly described when we can utilize such knowledge during the project. In this study, we conduct a practice that we collect collecting ominous signs in the ML projects relating to existing best practices as bad smells and report the results.

## 1 はじめに

社会の様々な領域で機械学習 (Machine Learning: ML) を用いた ML サービスシステムの開発プロジェクト (ML プロジェクト) が行われており、プロジェクトの実践を通して、別のプロジェクトでも適用可能な経験が得られている。一方、ML サービスシステムの開発では、従来のシステム開発以上に、開発者以外のプロジェクト関係者が要求からテスト、サービスイ

ン後の運用までに関わる必要がある。よって、再利用可能なノウハウや経験を活用し、開発をどのように効果的に進めていくかについてプロジェクト関係者間で共通理解を得ることはプロジェクトを効果的に実施する上で有効であると考えられている。

再利用可能と思われるプロジェクトでの経験はベストプラクティスやパターンと呼ばれる形式の知識として整備される。これらには典型的な課題に対する解決策が記載されていることが多い。このような解決策の中には、課題が明確になってから適用可能なものもあるが、予防的な解決策の場合、課題につながる兆候を検知した上で解決策を適用するものもある。通常のソフトウェアサービスシステムの開発では、課題につながる兆候は仕様書やプログラムコード中に埋め込まれていることが多い。実装やテストの工程に入る前に、そういった兆候を検出することができれば解決策を適用することができる。一方、ML サービスシステムの開発は、ML を使って解きたい予測問題や効果測定の指標に関する議論や、訓練に利用するデータの収

\* Collecting Bad Smells of Machine Learning Service System Projects

This is an unrefereed paper. Copyrights belong to the Authors.

Hironori Takeuchi, 武蔵大学, Musashi University.

Kota Imazaki, 情報処理推進機構, Information-technology Promotion Agency.

Yosuke Motohashi, 日本電気株式会社, NEC Corporation.

Noriyoshi Kuno, 三菱電機株式会社, Mitsubishi Electric Corporation.

Takuo Doi, デジタルアスリート株式会社, Digital Athlete Inc..

集や受け渡しといった活動の中に課題につながる兆候が含まれていることもある。こういった兆候は文書などの形で明示されることは少ないため、後から検知することは難しい。またそういった活動の場にデータサイエンティストがいない場合も多い。そのため課題につながる兆候があることにプロジェクト関係者が気がつかず、知識として共有された解決策を適用する機会を失うことがある。結果として、ML プロジェクトに関して再利用可能な知識を整備しても、プロジェクト実践の場で有効に活用できない状況が生じている。

そこで、本研究では ML サービスシステムの開発において、データサイエンティスト以外のプロジェクト担当者であっても利用できる形式で知見を整備することを考える。具体的には、ML プロジェクトにおける課題につながる兆候を危険な匂いとし、それを解決策と結びつけた形式で知見を整備する。すでに、収集されたベストプラクティスを解決策をもとに、それに関連する危険な匂いを実務家より収集し、知見として整備する。そして、収集の実践をワークショップを通して行い、危険な匂いを収集ができることを確認する。また、整備した知見やその活用方法についても考察する。

本論文の構成は以下のとおりである。2 節で関連研究を述べた後、3 節で研究対象と研究仮説について述べる。そして、4 節で提案手法を説明し、5 節で実践例とその結果を述べる。6 節で考察をした後、7 節でまとめを行う。

## 2 関連研究

文献調査や実プロジェクトへのヒアリングなどを通して、機械学習を活用したシステムを開発する ML プロジェクトにはソフトウェア工学上の課題が多くあることが示されている [4] [6]。課題解決の 1 つとして、ML プロジェクトを効果的に実施するための知識の外在化が重要であり、例えば、開発プロセスに関しては、汎用的なワークフローが示されている [1]。また、ML プロジェクトにおけるデータサイエンティストの役割についても [3] で体系化されている。そして、これら組み合わせ ML プロジェクトにおける各活動と、プロジェクト関係者およびプロジェクトのゴールを関

連づけたプロジェクトモデルも提案されている [8]。

一方、プロジェクトを進める上での知見としてベストプラクティスがある。ML プロジェクトについても文献調査を通してベストプラクティスを収集・整理し、それらの採用度合いや重要性を調査する試みがある [7]。また、システムアーキテクチャに関する知識として、ソフトウェアパターンがある [17]。機械学習を用いたシステムの開発に対してもパターンの検討が始まっており、体系的な文献収集を通じた ML サービスシステムのアーキテクチャ・デザインパターンの整備が進められ [12] [13]、プロジェクト実践におけるパターンの利用状況も調査されている [11]。

## 3 研究対象と研究仮説

### 3.1 ML サービスシステム

機械学習 (ML) は、説明変数 (特徴) を入力として目的変数 (結果) を予測する技術である。この時、説明変数と目的変数のペアからなるデータを訓練データとして集め、予測に用いるモデルを求める。つまり、ML サービスシステムは訓練データを ML アルゴリズムに投入し訓練済みモデル (ML モデル) を作成する部分 (ML 訓練エンジン) と、ML モデルに対して入力データを投入し、予測結果を出力として得る部分 (ML ランタイムエンジン) に分かれる。ML サービスシステムを開発する場合、対象業務のデータから訓練データを準備し、そこから訓練によって ML モデルを作るプロセスと、ML モデルを利用する ML ランタイムエンジンを使った業務アプリケーション (ML アプリケーション) を開発し、それをユーザーが利用するプロセスがある。これらの要素をエンタープライズアーキテクチャ (Enterprise Architecture: EA) の代表的なモデリング言語である ArchiMate [9] で表現すると図 1 となる。

### 3.2 ML プロジェクトにおけるベストプラクティス

ML サービスシステムに関するベストプラクティス [7] やアーキテクチャ・デザインパターン [12] [13] は文献収集に基づき整備されている。これらは、様々な粒度で書かれた文献からソフトウェア開発および機械学習に精通した専門家によって整備されている。また、

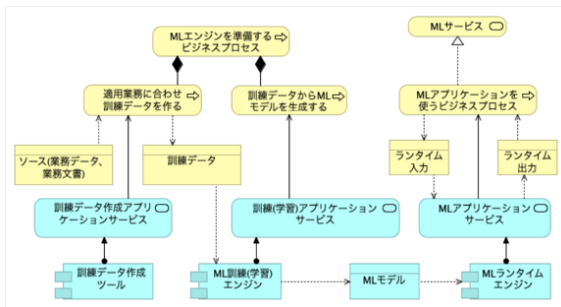


図1 ArchiMate で表現した ML サービスシステム

プロジェクト実践で得られた知見は、それが文献として公開されない限り、ベストプラクティスやパターンとして組織外にわたって整備・共有されることは難しい。様々な ML サービスシステムの開発を実施している大きな開発組織がそこで得られた知見をパターンとして公開している [5] が、パターンを作成する過程は明確にされていない。また、設計や実装以外にもプロジェクトの運営に関する知見が考えられるが、ML プロジェクトについて、こういった知見は特に上流工程を中心に十分には収集・整理が行われていない。そこで、ML プロジェクトの上流工程が詳細化されている参照モデルを利用し、プロジェクト実践から得られた様々な解決策をベストプラクティスとして収集し、パターンなどの知識として体系化する方法が提案されている [15]。図2に、その全体像を示す。このベ

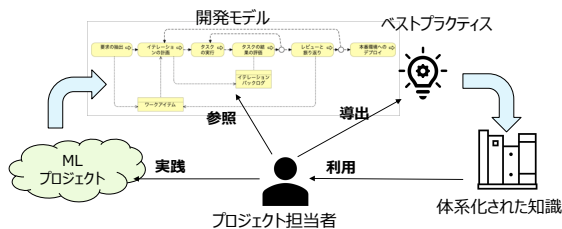


図2 モデル参照に基づいたベストプラクティスの収集

トプラクティスの収集および知識の体系化の手法は以下のステップからなる。

1. ML プロジェクトの実践に共通する開発の活動を抽出し、開発モデルを作成する

2. 開発モデルを参照し、プロジェクト実践で得られたベストプラクティスを収集する
3. 収集されたベストプラクティスに記載された解決策をもとにパターンを整備する

本手法をもとにプロジェクト担当者より収集したベストプラクティスの一部を以下に示す。

- システムの評価指標の検討においては、精度と同時に信頼性、安全性、公平性といった特性に関しても考える
- 計算資源についてプロジェクト内で確保する、もしくは合意をする
- データが収集される過程や方法を確認する

### 3.3 研究仮説

前節で述べたベストプラクティスの収集では、プロジェクト担当者は自身のプロジェクト実践で得られたベストプラクティスを提示する。その内容は、「開発におけるステージ」と「そのステージにおいて実施すべきこと」といった要素から構成されており、主に解決策が記載されている。このようなベストプラクティスにはガイドライン的なものもあり、それらはプロジェクト実施の際に常に適用されるべきものである。一方、何らかの技術的な課題に対する解決策が書かれていた場合、その解決策を適用する課題がどういった場面で発生するのかはベストプラクティス中に明記されていないことが多い。そのため、そういった解決策のみが記載されたベストプラクティスの活用はプロジェクト担当者の経験やスキルに依存することになる。結果、十分な経験のない実務家が多いプロジェクトでは蓄積された知見が有効に活用されないという状況が生じている。

そこで、本研究では以下を取り組むべき研究課題 (RQ) とする。

**RQ** ML プロジェクトの課題をプロジェクト担当者の経験に依存することなく解決することができる知見を整備することができるか

この RQ に対して、本研究では既存のベストプラク

ティスをもとに ML プロジェクトにおける危険な兆候を危険な匂いとして実務家より収集し、それらをベストプラクティスと紐付け知見として整備する手法を検討する。そして、実践を通して、知見を整備できることを確認するとともに、整備された知見について考察する。

## 4 ML プロジェクトにおける危険な匂いの収集

### 4.1 ML プロジェクトにおける危険な匂い

サービスシステムの開発を「開発者の作業」「変換処理」「稼働するシステム」の三段階で表現すると、従来のソフトウェアサービスシステムの開発と ML サービスシステムの開発は図 3 のように表される。開発者の作業の状況において、稼働するシステムに

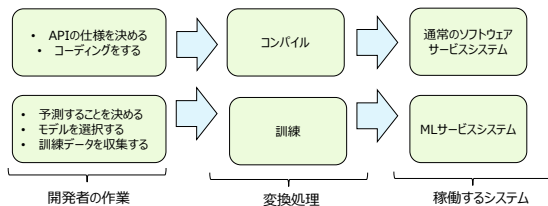


図 3 サービスシステムの開発

課題が生じることがある。開発者の作業によって埋め込まれる課題の発生につながる兆候は危険な匂い (bad smell) とも呼ばれる。危険な匂いのうち、コードに埋め込まれたものは code smell と呼ばれ、その検出が行われている。検出された code smell に対し、適切な解決策を適用することで稼働後に発生しうる課題を回避できるようになる。

ML プロジェクトにおいても、ML サービスシステムを構成する予測アプリケーションや訓練やその準備 (データ洗練など) に関するコードに埋められた code smell が整理されている [14]。一方、ML サービスシステムでは訓練に用いるデータにも潜在的な課題が埋まっている可能性があり、それが稼働するシステムの品質に大きな影響を与える。そのようなデータに埋まっている課題を data smell として整理する試みも実施されている [2]。これらの危険な匂いは、データ

やコードを直接観測することで検知することができる。そのため、訓練前やシステムの稼働前に解決策が適用できる可能性が高い。

一方、ML プロジェクトにおける開発者の作業には顧客との間の議論 (予測方法の決定やモデルの選択) や訓練データの収集や洗練といったものもある。この時、危険な匂いは「データの数が少ない」といった顧客から提供されたデータから観測できるもの以外に、議論の中で顧客の反応や発言といったような後の工程では検出が困難なものもある。そのような、成果物などから検出できないプロジェクト実施上の特徴を危険な匂いとして共有することができれば、プロジェクトの遂行中に担当者は危険な兆候を検知できると考えられる。

### 4.2 危険な匂いの収集

本研究では、問題解決プロセスモデル [10] に基づき危険な匂いを収集する。この問題解決プロセスモデルをサービスシステムの開発の文脈で表すと次のようになる。

1. 兆候に気づく
2. 起こり得る問題とその影響を把握する
3. 解決策を判断する
4. 解決策を適用する

この問題解決プロセスモデルを用いて知見を収集するためのフレームが表 1 として提案されている [16]。既存研究 [16] では表 1 の C に解決策を埋め、A, B

表 1 知見収集のフレーム

観測されるプロジェクトの状態	起こり得る課題	解決策
A	B	C

にそれぞれに対応する、観測されるプロジェクトの状態とそれによって引き起こされる可能性がある課題を埋める。

本研究では、このフレームの一部を活用して危険な匂いの収集を行い、知見を整備する。具体的には、以下に手順で行う。

1. 既存のベストプラクティスを解決策として実務家に提示する (C)
2. 各ベストプラクティスについて、どういった状況が観測されたら適用すべきかを危険な匂いとして洗い出し、対応づけを行う (A)
3. 2の他に危険な匂いがあれば適用する解決策とともに提示する (A, C)

これにより、既存のベストプラクティスに対して危険な匂いと紐付けし、知見として整備することができ、さらに新たな知見を危険な匂いと解決策の組み合わせという形式で得ることを目指す。

## 5 収集の実践例

本節では、提案手法を用いて危険な匂いを収集し、ベストプラクティスを対応づけ、知見として整備した実践結果について述べる。本実践では、2021年7月に実務家からのヒアリングを通して収集したMLプロジェクトのベストプラクティス[15]を利用する。そのベストプラクティスの内訳をSerbanらが文献から収集したベストプラクティス[7]と比較すると表2となる。プロジェクトの企画・設計段階におけるベストプラクティスが多く含まれ、顧客との議論の中に潜む危険な匂いを収集できることが期待される。

表2 本研究で用いるベストプラクティス

	計画・設計	データ	訓練	システム開発	デプロイ・運用	体制	ガバナンス	合計
本研究で使用するベストプラクティス	8	4	7	2	3	4	0	28
Serbanによるベストプラクティス	0	5	11	3	6	3	1	29

28のベストプラクティスをMuralと呼ばれるオンライン上の壁画ツールに付箋としてはり、ワークショップ<sup>†1</sup>をオンライン形式で実施した。ワークショップでは、MLプロジェクトを担当した経験のある6人の実務家がオンラインで議論をし、既存のベストプラクティスに対して、それを適用すべき危険な匂いをMural上に提示した。また同時に、新たな危険な匂いとそれに関連するベストプラクティスを提示した。Mural上で危険な匂いを収集し、ベストプラクティス

と紐付けた様子の一部を図4に示す。

あらかじめ提示した既存のベストプラクティスに対して収集された危険な匂い、そして、新たに収集された危険な匂い付きベストプラクティスの内訳を表3に示す。既存のベストプラクティスの53.6%に対し

表3 危険な匂いの収集結果

	計画・設計	データ	訓練	システム開発	デプロイ・運用	体制	合計
既存のベストプラクティス	8	4	7	2	3	4	28
既存のベストプラクティスの内危険な匂いが付与されたもの	5	3	5	0	0	2	15
新たに追加された危険な匂い付きベストプラクティス	7	1	1	0	0	0	9

て危険な匂いが収集され紐付けられた。MLプロジェクトでは、計画・設計から訓練までは開発側が顧客とのやりとりが多く発生するため、そこに危険な匂いが集中していることがわかる。また、新たに9個のベストプラクティスが危険な匂いとともに収集され、その多くがプロジェクトの計画・設計に関わるものであった。これらから、既存のベストプラクティスとして、計画・設計に関するものを多く含むものを準備したことが妥当であったと確認できた。また、既存のベストプラクティスを参照しながら該当する危険な匂いを検討することが、新たな知見を見出すことにつながる可能性があると考えられる。

Mural上に付箋として提示された、危険な匂いは、1つのベストプラクティスに複数紐付けられることもあった。また、ベストプラクティスを適用すべきプロジェクトの状況が具体的に記載されているものもあった。そこで、そのようなケースについては付箋に書かれた内容をまとめ、危険な匂いのタイトルを作り、それをベストプラクティスと紐づけて整理した。整理した結果の例を以下に示す。

- 危険な匂い：過度な期待と不明確な目的
  - － 顧客がMLに対して過度な期待を持っている
  - － 顧客が「AIが…」 「MLが…」とAIやMLを主語にした発言をする
- 解決策：ビジネス課題をMLの課題に落とし込み、合意を得る

†1 第5回機械学習工学研究会の企画セッション(2022年7月1日実施)

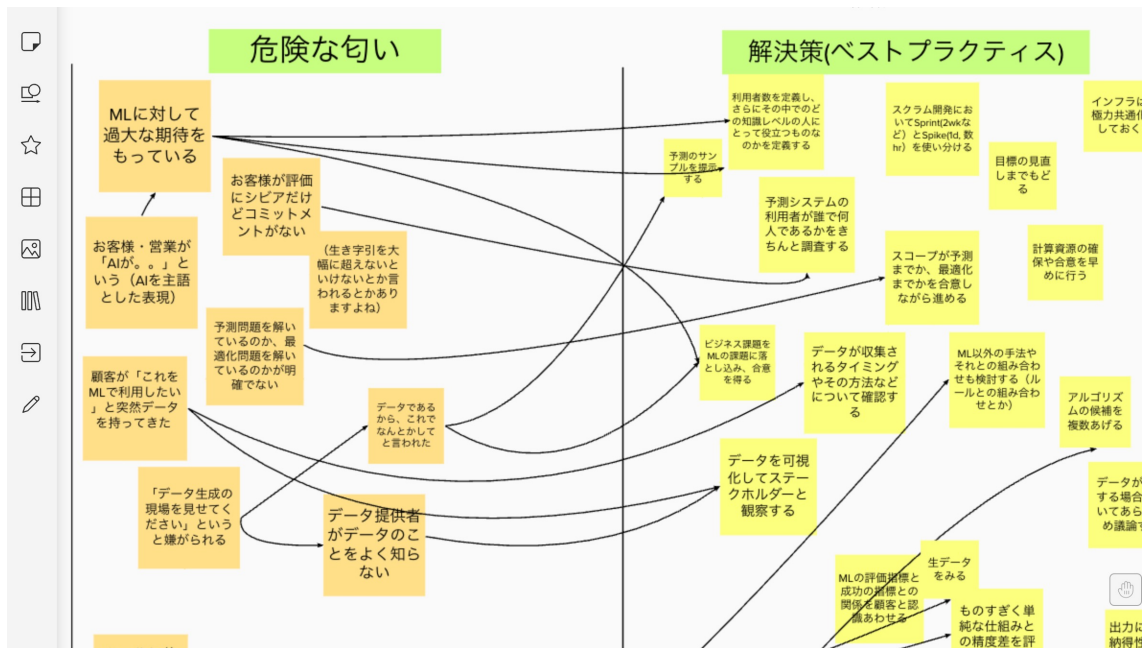


図4 危険な匂い (bad smell) を Mural 上で収集した結果 (一部)

## 6 考察

提案手法の実践により、ベストプラクティスに危険な匂いを紐づけた形で ML プロジェクトに関する知見を整備することができた。これにより、提案手法が RQ(ML プロジェクトの課題をプロジェクト担当者の経験に依存することなく解決することができる知見を整備することができるか)を解決する一手段となることが確認できた。また、既存のベストプラクティスについて、関連する危険な匂いを検討することで、新たな知見が想起されることも確認された。提案手法は、プロジェクト実践から網羅的に知見を整備していく上で有効な手段となり得ると期待される。

危険な匂いの例として「過度な期待と不明確な目的」が収集されたが、これは「ビジネス課題を ML の課題に落とし込み、合意を得る」というベストプラクティスだけでなく、「予測のサンプルの提示する」というベストプラクティスにも紐付けられた。このことから、危険な匂いとベストプラクティスは多対多 (M : N) の関係にあることがわかる。また、共通する危険な匂いを通してベストプラクティス同士の関連

性を調べることも可能となると期待できる。

本研究では、危険な匂いとベストプラクティスの組み合わせで ML プロジェクトに関する知見を表現した。これは、ソフトウェア開発パターンにおけるアンチパターンと捉えることができる。今回収集された 24 の危険な匂いとベストプラクティスの組み合わせをアンチパターンとして整備することは重要であり、今後の課題である。一方、収集された危険な匂いにはプロジェクトの具体的な状況を記載したものもあった。今回の収集では、それらを抽象化した記述を危険な匂いのタイトルとした。しかし、危険な匂いを抽象化した記述で表現することは、経験の浅いプロジェクト担当者にとって具体的な状況が想像できない、また、自身の経験と適切に結びつけられない可能性がある。これは、プロジェクト実施時に担当者が危険な匂いに相当する事象に遭遇しても気がつかないことを意味する。収集した危険な匂いをプロジェクト実践で役立てるため、どのような記載の粒度でモデル化するのが適切かどうかの調査が必要であり、これも今後の課題である。

## 7 まとめ

本研究では ML 技術を活用したサービスシステムの開発する ML プロジェクトを対象とした。近年、数多くの ML プロジェクトが行われており、プロジェクトを実施する上での様々な課題に対する解決策がベストプラクティスやパターンといった形の知識として蓄積されつつある。しかしながら、整備されたベストプラクティスがどういった状況の元で利用すべきかが不明確であるため、これらの蓄積された知識の活用はプロジェクト担当者の経験やスキルに依存する状態にあった。

そこで本研究では、ML プロジェクトについて課題を引き起こす可能性がある状況を危険な兆候としてとらえ、既存のベストプラクティスをもとに収集する手法を提案し、実践した。結果、28 の既存のベストプラクティスのうち 15 について危険な匂いを紐づけることができ、あらたに 9 組の危険な匂いとベストプラクティスの組み合わせを収集することができた。実践を通して得られた 24 組の危険な匂いとベストプラクティスの組み合わせをアンチパターンとして整備することや、危険な匂いをプロジェクト担当者にとってどのような粒度で記載すべきかの調査と検討が今後の課題である。

## 謝辞

本研究の一部は、JSPS 科研費 (JP19K20416) および JST 未来社会創造事業「機械学習を用いたシステムの高品質化・実用化を加速する”Engineerable AI”技術の開発」(JPMJMI20B8) の支援を受け実施した。

## 参考文献

- [1] Amershi, S., Begel, A., Bird, C., Deliner, R., Gall, H., Kamar, E., Nushi, N. N. B., and Zimmermann, T.: Software Engineering for Machine Learning: A Case Study, *Proceedings of the 41st International Conference on Software Engineering*, 2019, pp. 291 – 300.
- [2] Foidl, H., Felderer, M., and Ramler, R.: Data Smells: Categories, Causes and Consequences, and Detection of Suspicious Data in AI-based Systems, *Proceedings of the IEEE/ACM 1st International Conference on AI Engineering - Software Engineering (CAIN)*, 2022, pp. 229 – 239.
- [3] Kim, M., Zimmermann, T., DeLine, R., and Begel, A.: The Emerging Role of Data Scientists on Software Development Teams, *Proceedings of the 38th International Conference on Software Engineering*, 2016, pp. 96–107.
- [4] Kumeno, F.: Software Engineering Challenges for Machine Learning Applications: A Literature Review, *Intelligent Decision Technologies*, Vol. 13(2019), pp. 463 – 476.
- [5] Lakshmanan, V., Robinson, S., and Mann, M.: *Machine Learning Design Patterns: Solutions to Common Challenges in Data Preparation, Model Building, and MLOps*, O’Reilly, 2020.
- [6] Lwakatere, L. E., Raj, A., Bosch, J., Olsson, H. H., and Crnkovic, I.: A Taxonomy of Software Engineering Challenges for Machine Learning Systems: An Empirical Investigation, *Proceedings of the 20th International Conference on Agile Software Development (XP)*, 2019, pp. 227 – 243.
- [7] Serban, A., van der Blom, K., Hoos, H., and Visser, J.: Adoption and Effects of Software Engineering Best Practices in Machine Learning, *Proceedings of the ACM / IEEE International Symposium on Empirical Software Engineering and Measurement*, 2020, pp. 3:1 – 3:12.
- [8] Takeuchi, H. and Yamamoto, S.: AI Service System Development Using Enterprise Architecture Modeling, *Proceedings of the 23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (Procedia Computer Science vol. 159)*, 2019, pp. 923 – 932.
- [9] The Open Group: *ArchiMate 3.1 – A Pocket Guide*, Van Haren Publishing, 2019.
- [10] Tichy, N. M. and Bennis, W. G.: Making judgment calls. The ultimate act of leadership, *Harvard Business Review*, Vol. 85, No. 10(2007), pp. 94 – 102.
- [11] Washizaki, H., Khomh, F., Guéhéneuc, Y.-G., Takeuchi, H., Natori, N., Doi, T., and Okuda, S.: Software-Engineering Design Patterns for Machine Learning Applications, *IEEE Computer*, Vol. 55, No. 3(2022), pp. 30 – 39.
- [12] Washizaki, H., Khomh, F., Guéhéneuc, Y.-G., Takeuchi, H., Okuda, S., Natori, N., and Shioura, N.: Software Engineering Patterns for Machine Learning Applications (SEP4MLA) - Part 2, *Proceedings of the 27th Conference on Pattern Languages of Programs (PLoP 2020)*, 2020.
- [13] Washizaki, H., Uchida, H., Khomh, F., and Guéhéneuc, Y.-G.: Software Engineering Patterns for Machine Learning Applications (SEP4MLA), *Proceedings of the 9th Asian Conference on Pattern Languages of Programs (AsianPLoP 2020)*, 2020.
- [14] Zhang, H., Cruz, L., and van Deursen, A.: Code Smells for Machine Learning Applications, *Proceedings of the IEEE/ACM 1st International Conference on AI Engineering - Software Engineering (CAIN)*, 2022, pp. 229 – 239.

*ence on AI Engineering - Software Engineering (CAIN)*, 2022, pp. 217 – 228.

- [15] 竹内広宜, 今崎耕太, 久野倫義, 土肥拓生, 本橋洋介: 参照モデルを用いた機械学習プロジェクトの知見収集とパターンの作成, 信学技報 *KBSE2022-1*, 2022, pp. 8 – 13.
- [16] 竹内広宜, 小形真平, 海谷治彦, 中川博之: 意思決定

プロセスモデルを活用した機械学習プロジェクトに関する知見の体系化, 信学技報 *SC2022-5*, 2022, pp. 31 – 36.

- [17] 深澤良彰 監修, 鷺崎弘宜, 丸山勝久, 山本里枝子, 久保淳人 著: ソフトウェアパターン:パターン指向の実践ソフトウェア開発, 近代現代社, 2007.