



Онтология методов машинного обучения при управлении проектами

© 2026, Т.М. Мурзагалеев^{1,2}✉, Н.А. Жукова^{1,3}

¹ Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» (СПбГЭТУ «ЛЭТИ»), Санкт-Петербург, Россия

² ООО «РН-Проектирование Добыча», Томск, Россия

³ Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН), Санкт-Петербург, Россия

Аннотация

Применение методов машинного обучения в области управления проектами потенциально обеспечивает ряд преимуществ, которые могут позволить улучшить управление. Для эффективного применения методов машинного обучения используются наиболее показательные метрики с учётом разнообразия наборов данных, для которых требуется структурирование и систематизация элементов управления проектами. Целью данного исследования является разработка онтологии, позволяющей обеспечить необходимую информацию о применении методов машинного обучения в управлении проектами для прогнозирования трудозатрат и сроков выполнения проектов. Выполнен обзор методов машинного обучения, их метрик и наборов данных из научных публикаций, разработана и описана структура онтологии в области применения методов машинного обучения в управлении проектами для прогнозирования трудозатрат и сроков проектов, описана схема информационного взаимодействия пользователя с онтологией. Результаты проведённого исследования могут позволить упростить процесс поиска необходимых методов машинного обучения для их использования при управлении проектами на основе новых или общедоступных наборов данных.

Ключевые слова: онтология, управление проектами, прогнозирование срока проекта, трудозатраты проекта, методы машинного обучения, набор данных.

Цитирование: Мурзагалеев Т.М., Жукова Н.А. Онтология методов машинного обучения при управлении проектами. *Онтология проектирования*. 2026. Т.16, №2(60). С.300-313. DOI: 10.18287/2223-9537-2026-16-2-300-313.

Финансирование: исследование выполнено при поддержке государственного бюджета, номер проекта FFZF-2025-0019.

Вклад авторов: Мурзагалеев Т.М. – подбор и анализ источников, разработка онтологии; Жукова Н.А. – разработка структуры статьи, научное руководство, итоговые выводы.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Введение

Основными проблемами при управлении проектами (УП) в любой отрасли являются малоэффективное использование ресурсов, превышение ожидаемой продолжительности выполнения задачи и перерасход ресурсов, а также отсутствие непрерывного мониторинга процесса выполнения проекта [1, 2]. Точная оценка сроков и затрат проектов позволяет разрабатывать планы выполнения работ, выявлять и преодолевать потенциальные задержки и конфликты, а также обеспечивать баланс между стоимостью и качеством [3, 4].

Применение прогнозирующих методов машинного обучения (ММО) для оценки сроков и затрат при разработке проектов считается высокоэффективным способом устранения неопределённости, а также автоматизации решения повторяющихся задач, составления прогнозов и предоставления аналитических данных для эффективного принятия решений [5, 6].

Успешность применения ММО в УП зависит от наличия актуальных и высококачественных исходных данных. В качестве характеристик наборов данных (НД), используемых для построения модели, рассматриваются: размер НД, наличие выбросов, категориальных признаков, пропущенных значений и др. Разнообразие и сложность проекта в сочетании с присутствующими рисками затрудняют разработку модели, позволяющей точно оценить трудоёмкость проекта. Поэтому используется ряд подходов и методов, адаптированных к конкретным типам проектов внутри отдельных отраслей [7]. Для эффективного применения ММО с учётом разнообразия НД требуется структурирование и систематизация ММО, метрик и НД в исследуемой предметной области (ПрО).

Целью данного исследования является разработка онтологии, позволяющей обеспечить интеграцию информации о применении ММО в УП для прогнозирования трудозатрат и сроков выполнения проектов.

1 Проектирование онтологии

Онтология является формой целостного представления системы знаний конкретной ПрО с помощью построения иерархии понятий (классов) и структуры отношений между ними [8]. В данной статье представлена онтология использования ММО для прогнозирования трудозатрат проектов при УП, как инструмент для структурирования и классификации данных [9].

В исследовании используется редактор *Protégé*, с помощью которого разрабатывается онтология. Запись семантических моделей осуществляется на языках *RDF/RDFS* и *OWL*. *RDF/RDFS* позволяют записывать простейшие факты об объектах, классах и свойствах. *RDF* предоставляет средства для записи триплетов: субъект–предикат–объект. *OWL*-онтология может включать описания классов, свойств и их экземпляров и определяет, как получать логические следствия, которые не присутствуют непосредственно в онтологии, но могут быть выведены из существующих с использованием семантики [10].

При построении онтологии выполняется следующая последовательность действий.

1) **определение классов** (*Classes*): типов объектов, их характеристик, комплексов понятий с их участием. В состав онтологии включены следующие классы: ММО (*Method_ML*), метрика ММО (*Metric_ML*), эксперимент (*Experiment*). Фундаментальным таксономическим конструктором для классов является отношение «быть подклассом» (*rdfs:subClassOf*), описывающее связь частного класса с более общим [11]. В разработанной онтологии классы *Experiment*, *Method_ML* и *Metric_ML* являются подклассами суперкласса *owl:Thing* (см. рисунок 1).

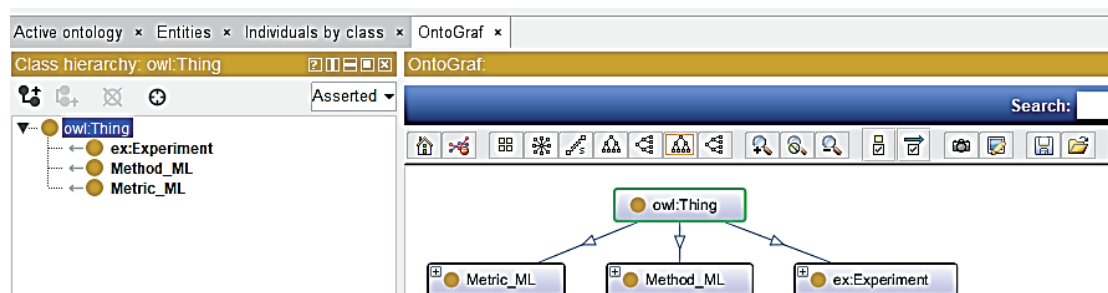


Рисунок 1 – Представление иерархии классов в редакторе *Protégé*

2) **определение отношений** (*Object properties*) (см. рисунок 2). Отношения определяют семантические связи между экземплярами (индивидами) различных классов. При связи двух сущностей через *Object properties* одно из них является объектом, а другое субъектом. Например, класс ММО имеет связь с классом метрика ММО, в онтологии – это отношение *Has_metric*, в данной группе класс ММО определяется объектом, а класс метрика ММО – субъектом с помощью описания отношения в секциях *Domains* и *Ranges*. Отношение означает, что любой метод имеет как минимум одну метрику. Отношения, имеющиеся в онтологии, представлены в таблице 1.

3) **формирование** схемы онтологии как связанного комплекса компонентов. Схема определяется структурой подчинения классов и отношениями между классами.

4) **дополнение** онтологии экземплярами классов (*Individuals*), являющихся основными нижеуровневыми компонентами онтологии (см. рисунок 3), а также данными, имеющими физический смысл.

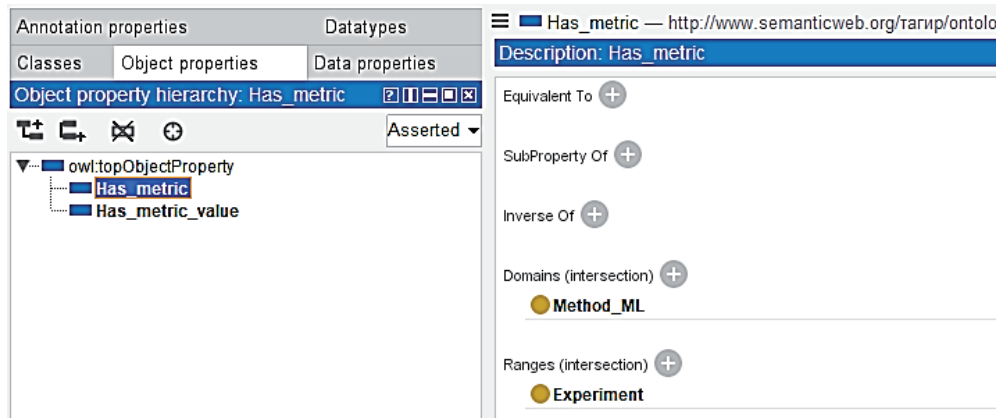


Рисунок 2 – Определение *Object properties* в редакторе *Protégé* на примере *Has_metric*

Таблица 1 – Связь классов в онтологии через *Object properties*

Отношения в онтологии	Описание отношения	Классы, участвующие в отношении
<i>Has_metric</i>	Связь указывает на наличие эксперимента со значением метрики у метода, причём экспериментов может быть несколько у метода	<i>Method_ML</i> , <i>Experiment</i>
<i>Has_metric_value</i>	Связь указывает на тип метрики в эксперименте со значением этой метрики, причём в нескольких экспериментах может быть использован один и тот же тип метрики	<i>Experiment</i> , <i>Metric_ML</i>

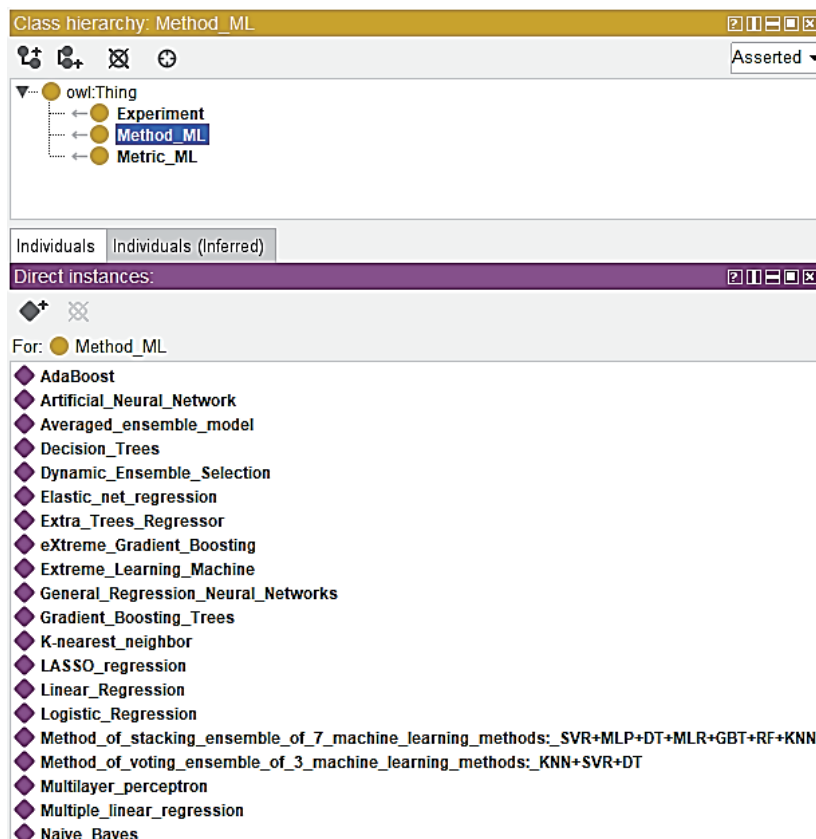


Рисунок 3 – *Individuals* класса *Method_ML* в редакторе *Protégé*

5) **определение свойств данных (Data properties) Individuals**, связывая их с конкретными значениями, такими как строки, числа, даты или логические значения. Они не соединяют объекты между собой, а определяют их характеристики, см., например, рисунок 4: индивид *_1* класса *Experiment* имеет свойства НД (*Data_set*), ссылка на источник (*Source_DOI*), значение метрики (*Value*). Свойства данных представлены в таблице 2.

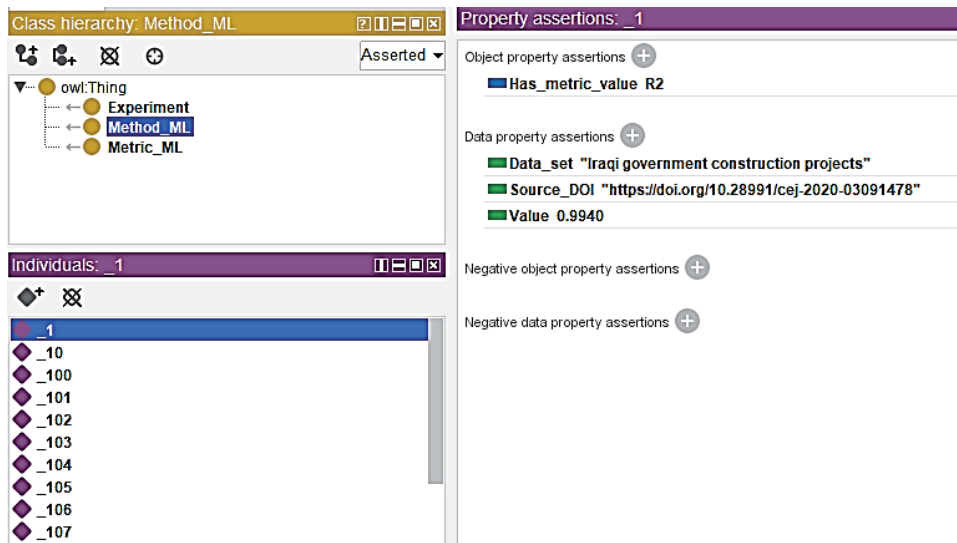


Рисунок 4 – Data properties класса Experiment в редакторе Protégé

Таблица 2 – Data properties индивидов классов в онтологии

Наименование свойства данных в онтологии	Описание свойства данных	Класс, для которого задано свойство данных
Value	Полученное значение метрики в эксперименте	Experiment
Data_set	Набор данных, на котором выполнялся эксперимент	
Source_DOI	Ссылка на работу, в которой представлен эксперимент	

Описание ПрО представлено на рисунке 5 в виде графа.

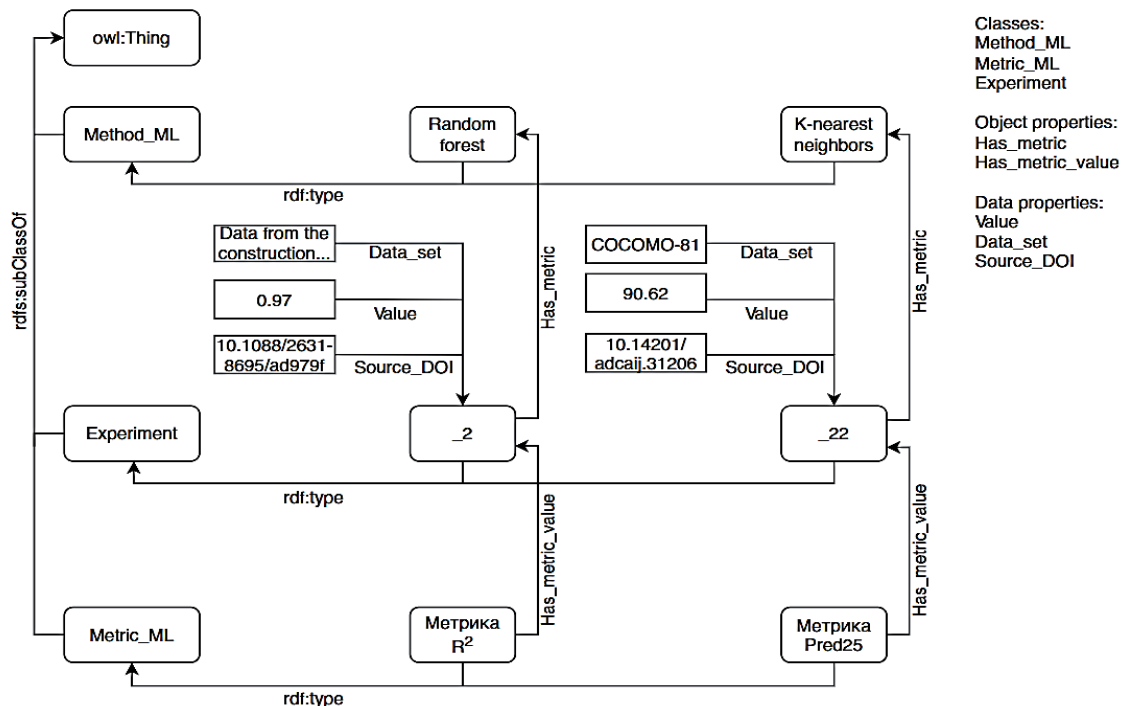


Рисунок 5 – Фрагмент графа, описывающего классы, экземпляры и их отношения в предметной области

Связи объекта и субъекта обозначены линией с направлением стрелки в сторону объекта. Линией с направлением стрелки обозначены отношения между индивидами, на которой направление стрелки обозначает, какой индивид является объектом, а какой – субъектом.

2 Наполнение онтологии

Для сбора применяемых ММО в УП при прогнозировании трудозатрат и сроков выполнения проектов были проанализированы публикации за период с 2017 по 2025 год.

В [12] для оценки стоимости и продолжительности строительства применён алгоритм оптимизации роя частиц. Для построения модели использована серия из 60 проектов, осуществленных в период с 2008 по 2016 года. Основными метриками в данном исследовании были корень из среднеквадратичной ошибки ($RMSE$) и коэффициент детерминации (R^2).

В [13] ММО применены для прогнозирования продолжительности строительства. При оценке ММО использовались $RMSE$, R^2 и средняя абсолютная погрешность (MAE). Например, модель случайного леса (*Random Forest, RF*) достигла значений $RMSE$ в 74 дня и R^2 0.97.

В [7] рассматривался потенциал ММО для оценки временных задержек в проектах морского строительства. Используются среднеквадратичная ошибка (MSE), MAE , $RMSE$, индекс разброса (SCI) и коэффициент корреляции (R). НД включал 30 проектов из государственных и частных секторов ряда стран. Искусственная нейронная сеть общей регрессии показала высокую прогностическую эффективность по сравнению с другими моделями.

В [14] исследование проведено на примере данных о строительных работах в порту. Среди использованных моделей наилучшие результаты в прогнозировании сроков выполнения проектов показала модель *Silverkite*.

В [15] использованы четыре различных ММО. В НД было 722 вектора. Сравнение результатов прогнозных моделей показало, что искусственная нейронная сеть и градиентный бустинг деревьев (*Gradient Tree Boost, GBT*) дали лучшие результаты.

В [16] предложена модель множественной линейной регрессии для прогнозирования продолжительности строительства офисных зданий.

В [17] использованы общедоступные НД: *Albretch, Desharnais, COCOMO81, NASA, Kemerer, China* и *Kitchenham*. Эффективность методов измерялась с помощью показателей: средняя величина относительной погрешности ($MMRE$), прогнозы, которые находятся в пределах 0.25 от фактического целевого значения $PRED(0.25)$ и R^2 . Для НД *Albretch, NASA* лучшая модель – гребневая регрессия, для НД *Desharnais, COCOMO81, China, Kitchenham* – метод деревьев принятия решений, для НД *Kemerer* – многослойный перцептрон.

В [18] предложена усовершенствованная модель ансамбля с многоуровневой структурой *Stacked Ensemble*. Используются НД *COCOMO81*, содержащий данные по 63 программным проектам, методы *RF* и *GBT* для отбора признаков. Для прогнозирования трудозатрат разработки программных продуктов рассматривали различные ММО. Модель *Stacked Ensemble* показала высокие значения точности прогнозирования $PRED(0.25)$.

В [19] приведено сравнение трёх моделей нечёткой логики для прогнозирования затрат на разработку программного обеспечения: *Mamdani, Sugeno* с постоянным выходом и *Sugeno* с линейным выходом. Используются данные промышленных проектов из НД *ISBSG* с 5000 проектами. Эти модели нечёткой логики были сравнены с моделью множественной линейной регрессии и искусственной нейронной сетью с прямой передачей.

В [20] оценены различные модели штабелирования. Для оценки использованы НД *Albrecht, China, Desharnais, Kemerer, Kitchenham, Maxwell* и *COCOMO81*. Лучшие результаты были получены на модели штабелирования с использованием *RF*.

В [21] выполнено сравнение модели экстремальной обучающей машины (*ELM*) с регрессионными моделями: линейная регрессия, многослойный перцептрон, *k*-ближайших соседей, метод опорных векторов. *ELM* показал наилучшие результаты.

В [22] сравнивали восемь алгоритмов ММО на трёх НД: *SCRAIM*, *JIRA*, *Project Control*.

В [23] выполнена оценка трудозатрат для двух НД: *Finnish* и *Maxwell*. Первый этап состоял в нормализации данных, на втором этапе предложена улучшенная версия статического отбора ансамбля, основанная на генетическом алгоритме, на третьем этапе применён динамический отбор ансамбля. Используются метрики: симметричная средняя абсолютная процентная ошибка, средняя относительная ошибка, средняя абсолютная шкалированная ошибка, коэффициент Нэша-Сатклиффа, R^2 .

В [24] для анализа использования ММО для улучшения оценки производительности программного обеспечения использованы НД *ISBSG*, *NASA93*, *COCOMO*, *Maxwell* и *Desharnais*. Для НД с применением метода отбора признаков по коэффициентам корреляции линейная регрессия позволила получить $R^2=0.77$ для НД *ISBSG* и 1.0 для НД *NASA93*, *COCOMO*, *Maxwell* и *Desharnais*. Метод *RF* для НД *NASA93* и *Desharnais* показал R^2 равный 0.99.

В [25] для создания модели прогнозирования затрат применены различные ММО и метрика *MMRE*.

В [26] использован НД из семи признаков, состоявший из 2000 выборок данных, которые подавались в режиме реального времени. Результаты исследования показали, что метод дерева принятия решений даёт более точные результаты.

В [27] проанализированы восемь ММО на следующих НД: *Albrecht*, *COCOMO81*, *Desharnais*, *China*, *Finnish*, *Maxwell*, *Miyazaki*, *NASA18*, *NASA93*, *Telecom*, *Kitchenham* и *Kemerer*. Выявлено, что методы *RF* и *k*-ближайших соседей превзошли остальные.

В [28] для оценки трудозатрат использованы НД *Albrecht*, *China*, *Desharnais*, *Maxwell*. Алгоритм *RF* обеспечил лучшие результаты по метрикам *MAE*, *MSE*, *RMSE* и R^2 .

В [29] разработана модель штабелирования, основанная на поправочных коэффициентах оптимизации, путём интеграции семи статистических методов и ММО. Точность оценки методов оценивалась с использованием статистических тестов и показателей эффективности на основе четырёх НД, с метриками *SSE*, *MAE*, *RMSE*, *MBRE*, *MIBRE*, *MdMRE* и *PRED(0,25)*. Ансамблевая модель показала лучшие результаты.

В [6] выполнено усреднение по совокупности трёх наиболее эффективных ММО, а в [5] описана оптимизированная модель *RF* путём изменения значений её ключевых параметров.

На основе проведённого обзора ММО, НД и метрик, отбора методов и данных с экспериментальными результатами, которые могут быть полезными в практических целях, наполнена онтология в *Protégé*. Общая графовая схема онтологии представлена на рисунке 6, а её фрагмент – на рисунке 7. На рисунках каждый индивид класса *Experiment* обозначен с использованием сквозной нумерации (например, «_1», «_2», «_11»), названия методов соответствуют их общераспространённым названиям (например, «Random_forest», «Decision_trees»), в названиях метрик – общераспространённые обозначения. Фиолетовые сплошные линии со стрелками на схеме указывают на связь индивида с классом (стрелка по направлению к индивиду, например, *Method_ML* и *Random_forest*, *Metric_ML* и *Pred25*). Прерывистые оранжевые линии на схеме указывают на отношения между классами, стрелка указывает направление от субъекта к объекту, на схеме – это связи между *Method_ML* и *Experiment*, а также *Experiment* и *Metric_ML*. Синяя линия со стрелкой на схеме – это связь классов с суперклассом *owl:Thing*, который является родительским классом по отношению ко всем классам. Прерывистые серые и оранжевые линии на схеме указывают на связи между индивидами различных классов, заданные через отношения между классами. Например, см. рисунок 7, оранжевая линия между индивидом _64 класса *Experiment* и индивидом *Random_forest* класса *Meth-*

од *ML* отражает отношение *Has_metric*. Это интерпретируется как метод *RF* использован в эксперименте *_64*. Серая линия между индивидом *_64* класса *Experiment* и индивидом *R²* класса *Metric_ML* показывает отношение *Has_metric_value*, что интерпретируется как: в эксперименте *_64* для оценки применена метрика *R²*; при этом эксперимент *_64* в своих свойствах содержит значение метрики *R²*, ссылку на исследование и наименование НД. Направление стрелки указывает какая из сущностей является объектом, а какая – субъектом.

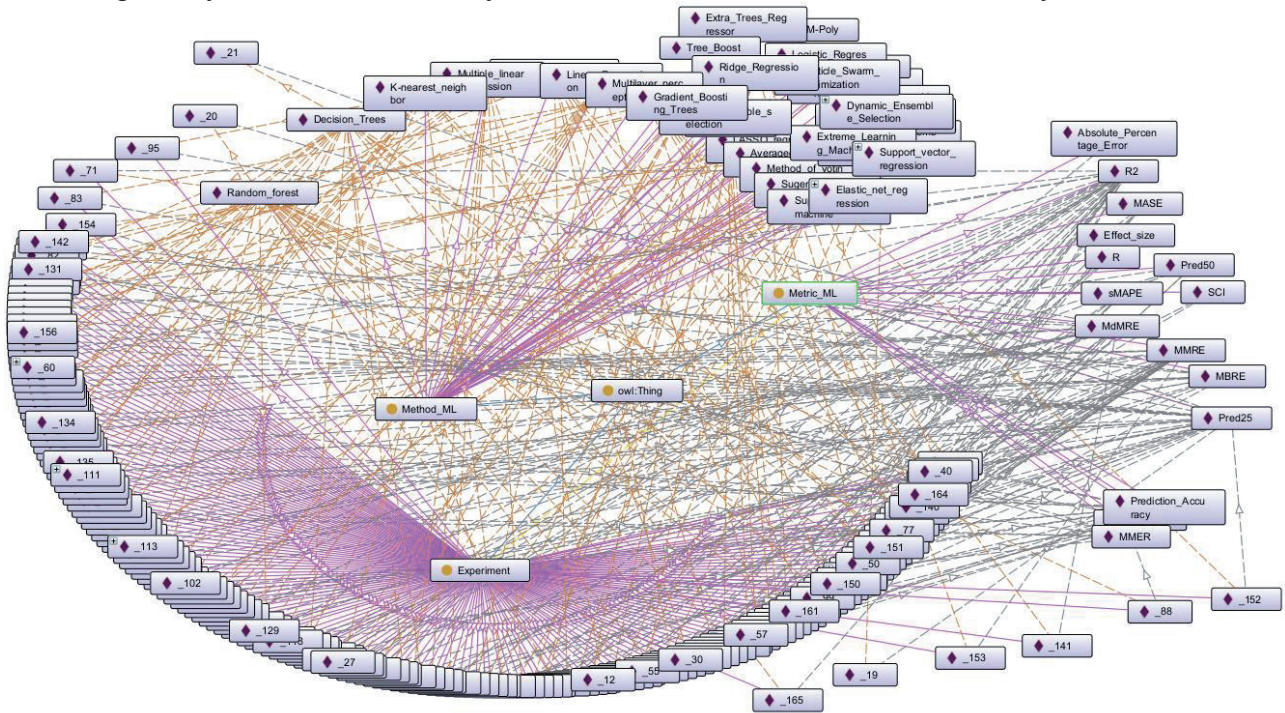


Рисунок 6 – Общая графовая схема онтологии в Protégé

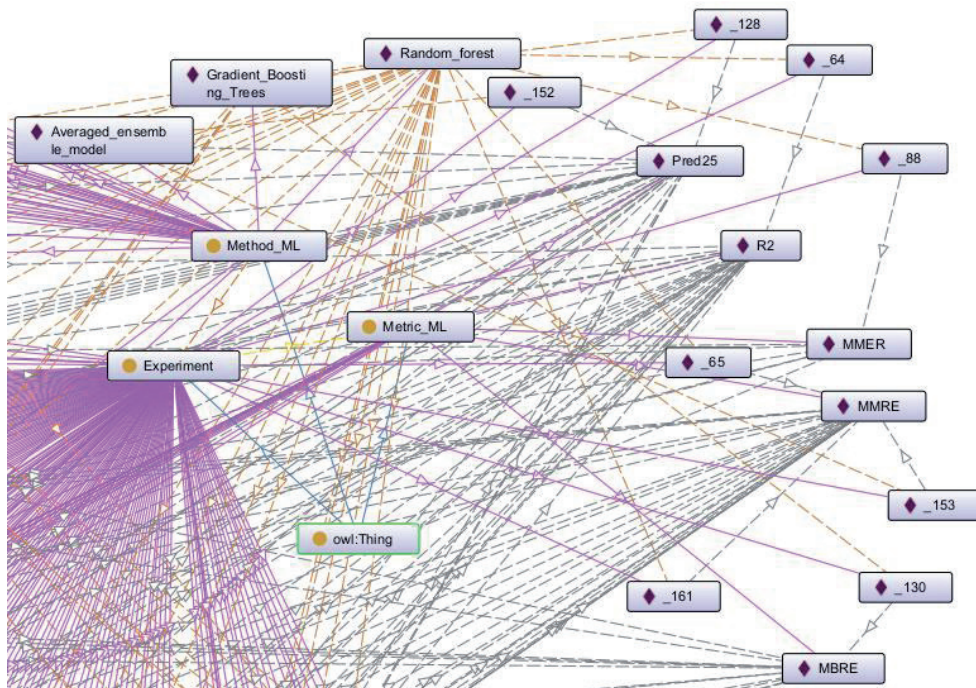


Рисунок 7 – Фрагмент графовой схемы онтологии в Protégé

Для возможности применения онтологии файл выложен в репозиторий *github*: https://github.com/TagirM/Ontology_methods_ML.git.

3 Использование онтологии

В данной онтологии можно с помощью *SPARQL*-запроса найти интересующую информацию, например, в виде следующих запросов.

Какие методы, с какими метриками и их значениями, НД и ссылками на исследования представлены в онтологии? Пример запроса приведён на рисунке 8. В выводе запроса в первом столбце выведены все методы (на рисунке показана только видимая часть), во втором столбце – эксперименты. В других столбцах в соответствии с запросом, приведённом в верхней части рисунка, указаны все наименования столбцов, стоящие в запросе после ключевого слова *WHERE*: *experiment, data_set, metric, value_str, reference*.

Какой НД использован, какие метрики применялись, какие значения были получены со ссылкой на исследования? Пример такого запроса представлен на рисунке 9. Для получения информации по определённому методу в запросе добавляется ключевое слово *FILTER*, после которого указывается нужный для вывода метод (на рисунке 9 сделан фильтр по методу *RF*).

Найти информацию с выводом всех экспериментов, в которых использована определённая метрика оценки результата работы ММО, и её значение находилось в заданном диапазоне. Пример запроса, где в качестве метрики оценки результата работы ММО использован R^2 со значением больше 0.90, с выводом используемого ММО, НД и ссылки на исследование, представлен на рисунке 10. Для получения данной информации с конкретной метрикой в запросе дважды указывается ключевое слово *FILTER*. После первого ключевого слова *FILTER* указывается необходимая для вывода метрика, после второго – диапазон величины метрики.

SPARQL query: ⏏ ⏏ ⏏

```

PREFIX db: <http://www.semanticweb.org/tagir/ontologies/2025/10/machine-learning-ontology#>

SELECT ?method ?experiment ?data_set ?metric(STR(?value) AS ?value_str) ?reference
WHERE {
?method rdf:type db:Method_ML.
?method db:Has_metric ?experiment.
?experiment db:Data_set ?data_set.
?experiment db:Has_metric_value ?metric.
?experiment db:Value ?value.
?experiment db:Source_DOI ?reference.
}

```

method	experiment	data_set	metric	value_str	reference
Omni-Ensemble_selection	_46	"Maxwell"	MASE	"0.26125"	"https://doi.org/10.1109/ACCESS.2017.1713111"
Logistic_Regression	_83	"Dataset on crowdsourcin	Pred25	"72.25"	"https://doi.org/10.3233/IDA-237361"
Omni-Ensemble_selection	_47	"Maxwell"	NSE	"0.95146"	"https://doi.org/10.1109/ACCESS.2017.1713111"
Averaged_ensemble_model	_152	"ISBSG"	Pred25	"69.44"	"https://doi.org/10.1016/j.jss.2017.05.001"
Logistic_Regression	_82	"Dataset on crowdsourcin	MdMRE	"0.049"	"https://doi.org/10.3233/IDA-237361"
Omni-Ensemble_selection	_44	"Maxwell"	R2	"0.98359"	"https://doi.org/10.1109/ACCESS.2017.1713111"
Random_forest	_71	"NASA93, Desharnais"	R2	"0.99"	"https://doi.org/10.18280/isi.28061"
Averaged_ensemble_model	_153	"ISBSG"	MMRE	"0.17"	"https://doi.org/10.1016/j.jss.2017.05.001"
Artificial_Neural_Network	_7	"Construction project data	R2	"0.944"	"https://doi.org/10.1080/13467581.2017.13467581"
Logistic_Regression	_81	"Dataset on crowdsourcin	MMRE	"0.005"	"https://doi.org/10.3233/IDA-237361"
Omni-Ensemble_selection	_45	"Maxwell"	MMRE	"0.47942"	"https://doi.org/10.1109/ACCESS.2017.1713111"
Averaged_ensemble_model	_154	"ISBSG"	MMER	"0.42"	"https://doi.org/10.1016/j.jss.2017.05.001"
Averaged_ensemble_model	_155	"ISBSG"	MBRE	"0.2"	"https://doi.org/10.1016/j.jss.2017.05.001"

Рисунок 8 – *SPARQL*-запрос в *Protégé* с выводом методов и экспериментов

SPARQL query:

```

PREFIX db: <http://www.semanticweb.org/taurip/ontologies/2025/10/machine-learning-ontology#>

SELECT ?method ?experiment ?data_set ?metric(STR(?value) AS ?value_str) ?reference
WHERE {
?method rdfs:type db:Method_ML.
?method db:Has_metric ?experiment.
?experiment db:Data_set ?data_set.
?experiment db:Has_metric_value ?metric.
?experiment db:Value ?value.
?experiment db:Source_DOI ?reference.
FILTER(?method = db:Random_forest)
}
    
```

method	experiment	data_set	metric	value_str	reference
Random_forest	_71	"NASA93, Desharnais"	R2	"0.99"	"https://doi.org/10.18280/isi.280602"
Random_forest	_2	"Data from the construction orga	R2	"0.97"	"https://doi.org/10.1088/2631-8695/ad979f"
Random_forest	_120	"Historical dataset from Software	Pred25	"1.0"	"https://doi.org/10.1002/smr.2611"
Random_forest	_121	"Historical dataset from Software	MdMRE	"0.013"	"https://doi.org/10.1002/smr.2611"
Random_forest	_122	"Historical dataset from Software	MBRE	"0.011"	"https://doi.org/10.1002/smr.2611"
Random_forest	_123	"Historical dataset from Software	MIBRE	"0.011"	"https://doi.org/10.1002/smr.2611"
Random_forest	_20	"COCOMO-81"	Pred25	"92.18"	"https://doi.org/10.14201/adcaij.31206"
Random_forest	_68	"Maxwell"	sMAPE	"18.274"	"https://doi.org/10.1109/ACCESS.2023.329"
Random_forest	_67	"Maxwell"	NSE	"0.88611"	"https://doi.org/10.1109/ACCESS.2023.329"
Random_forest	_66	"Maxwell"	MASE	"0.32081"	"https://doi.org/10.1109/ACCESS.2023.329"
Random_forest	_89	"NASA18"	MdMRE	"0.1140"	"https://doi.org/10.1016/j.jer.2023.100150"
Random_forest	_65	"Maxwell"	MMRE	"0.69642"	"https://doi.org/10.1109/ACCESS.2023.329"
Random_forest	_88	"NASA18"	MMER	"0.1748"	"https://doi.org/10.1016/j.jer.2023.100150"

Рисунок 9 – SPARQL-запрос в Protégé с выводом экспериментов при применении метода *Random_forest*

SPARQL query:

```

PREFIX db: <http://www.semanticweb.org/taurip/ontologies/2025/10/machine-learning-ontology#>

SELECT ?method ?experiment ?data_set ?metric(STR(?value) AS ?value_str) ?reference
WHERE {
?method rdfs:type db:Method_ML.
?method db:Has_metric ?experiment.
?experiment db:Data_set ?data_set.
?experiment db:Has_metric_value ?metric.
?experiment db:Value ?value.
?experiment db:Source_DOI ?reference.
FILTER(?metric = db:R2)
FILTER(?value > 0.9)
}
    
```

method	experiment	data_set	metric	value_str	reference
Gradient_Boosting_Trees	_8	"Construction project da	R2	"0.935"	"https://doi.org/10.1080/13467581.2023"
Random_forest	_71	"NASA93, Desharnais"	R2	"0.99"	"https://doi.org/10.18280/isi.280602"
Omni-Ensemble_selection	_44	"Maxwell"	R2	"0.98359"	"https://doi.org/10.1109/ACCESS.2023.3"
Linear_Regression	_102	"NASA18"	R2	"0.9972"	"https://doi.org/10.1016/j.jer.2023.10015"
Extra_Trees_Regressor	_41	"Project Control"	R2	"0.96"	"https://doi.org/10.1109/ACCESS.2023.3"
Stacking_ensemble_model_	_31	"China"	R2	"0.984"	"https://doi.org/10.3390/electronics1010"
eXtreme_Gradient_Boosting	_43	"Project Control"	R2	"0.96"	"https://doi.org/10.1109/ACCESS.2023.3"
Particle_Swarm_Optimizator	_1	"Iraqi government constr	R2	"0.9940"	"https://doi.org/10.28991/cej-2020-0309"
Linear_Regression	_69	"NASA93, Desharnais"	R2	"1.00"	"https://doi.org/10.18280/isi.280602"
eXtreme_Gradient_Boosting	_70	"NASA93, Desharnais"	R2	"0.99"	"https://doi.org/10.18280/isi.280602"
Dynamic_Ensemble_Selectio	_59	"Maxwell"	R2	"0.9522"	"https://doi.org/10.1109/ACCESS.2023.3"
Random_forest	_37	"Project Control"	R2	"0.96"	"https://doi.org/10.1109/ACCESS.2023.3"
Support_vector_regression	_9	"Construction project da	R2	"0.931"	"https://doi.org/10.1080/13467581.2023"

Рисунок 10 – SPARQL-запрос в Protégé с выводом экспериментов, в которых в качестве метрики оценки результата работы метода машинного обучения использован $R^2 > 0.90$

Заключение

Для описания ПрО «Прогнозирование трудозатрат и сроков выполнения задач при УП» построена онтология, при разработке которой выполнен анализ и систематизированы различные методы прогнозирования. Разработанная онтология является основой для структурирования исследований, проводимых в области применения ММО, извлечения необходимых пользователям данных и их контекстуализации. Применение результатов исследования может позволить упростить процесс поиска необходимых методов для их использования при прогнозировании трудозатрат и сроков выполнения проектов на основе новых или общедоступных НД.

Разработанная онтология имеет ограничения: не все существующие исследования занесены в базу знаний; не учтены параметры НД, такие как размер данных, количество признаков, форматы признаков (категориальные и числовые); ограниченное количество метрик.

СПИСОК ИСТОЧНИКОВ

- [1] **Prasad K.S.N., Vijaya Saradhi M.V.** Comprehensive project management framework using machine learning. *International Journal of Recent Technology and Engineering (IJRTE)*. 2019. Vol.8. Issue 2S3. DOI: 10.35940/ijrte.B1256.0782S319.
- [2] **Santos J.I., Pereda M., Ahedo V., Galán J.M.** Explainable machine learning for project management control. *Computers & Industrial Engineering*. 2023. Vol.180. 109261. DOI: 10.1016/j.cie.2023.109261.
- [3] **Forouzesht Nejad A.A., Arabikhan F., Aheleroff S.** Optimizing project time and cost prediction using a hybrid XGBoost and simulated annealing algorithm. *Machines*. 2024. №12(12). 867. DOI: 10.3390/machines12120867.
- [4] **Дударин П.В., Тронин В.Г., Святлов К.В.** Подход к оценке трудоёмкости задач в процессе разработки программного обеспечения на основе нейронных сетей. *Автоматизация процессов управления*. 2019. №3(57). С.65–72. DOI: 10.35752/1991-2927-2019-3-57-65-72.
- [5] **Abdelali Z., Mustapha H., Abdelwahed N.** Investigating the use of random forest in software effort estimation. *Procedia Computer Science*. 2019. Vol.148, P.343-352. DOI: 10.1016/j.procs.2019.01.042.
- [6] **Pospieszny P., Czarnacka-Chrobot B., Kobylinski A.** An effective approach for software project effort and duration estimation with machine learning algorithms. *Journal of Systems and Software*. 2018. Vol.137. P.184-196. DOI: 10.1016/j.jss.2017.11.066.
- [7] **Nassar A.H., Elbisy A.M.** A machine learning approach to predict time delays in marine construction projects. *Engineering, Technology & Applied Science Research*. 2024. Vol.14. №5. P.16125–16134. DOI: 10.48084/etasr.8173.
- [8] **Напольских Д.Л.** Представление экономической концепции кластера с помощью специализированной онтологии. *Программные продукты и системы*. 2024. Т.37. №2. С.155–163. DOI: 10.15827/0236-235X.142.155-163.
- [9] **Gruber T.R.** A translation approach to portable ontology specifications. Knowledge Acquisition, 1993.
- [10] **Исаева О.С.** Построение онтологии для систематизации характеристик сети Интернета вещей. *Онтология проектирования*. 2024. Т.14. №2(52). С.243-255. DOI: 10.18287/2223-9537-2024-14-2-243-255.
- [11] **Гусенков А.М., Бухараев Н.Р., Бирияльцев Е.В.** Построение онтологии предметной области на основе логической модели данных. *Электронные библиотеки*. 2020. №23(3). С.390-417. DOI: 10.26907/1562-5419-2020-23-3-390-417.
- [12] **Khalaf T.Z., Çağlar H., Çağlar A., Hanoon A.N.** Particle swarm optimization based approach for estimation of costs and duration of construction projects. *Civil Engineering Journal*. 2020. №6(2). P.384–401. DOI: 10.28991/cej-2020-03091478.
- [13] **Soreti M. Liben, Demiss A. Belachew, Walied A. Elsaigh.** Comparing advanced and traditional machine learning algorithms for construction duration prediction: a case study of Addis Ababa's public sector. *Engineering Research Express*. 2024. Vol. 6. № 4. DOI: 10.1088/2631-8695/ad979f.
- [14] **Sklias G., Gkelios S., Dimitriou D., Sartzetaki M., Chatzichristofis S.A.** Synergizing global and local strategies for dynamic project management: An Advanced Machine Learning-Enhanced Framework. *IEEE Access*. 2024. Vol.12. P.85955-85968. DOI: 10.1109/ACCESS.2024.3413890.
- [15] **Zhang S., Xuechun Li.** A comparative study of machine learning regression models for predicting construction duration. *Journal of Asian Architecture and Building Engineering*. 2023. №23(6). DOI: 10.1080/13467581.2023.2278887.

- [16] **Yeom D.J., Seo H.M., Kim Y.J., Cho C.S., Kim Y.** Development of an approximate construction duration prediction model during the project planning phase for general office buildings. *Journal of Civil Engineering and Management*. 2018. №24(3). P.238-253. DOI: 10.3846/jcem.2018.1646.
- [17] **Israr ur Rehmana, Zulfiqar Alib, Zahoor Jana.** An empirical analysis on software development efforts estimation in machine learning perspective. *Advances in Distributed Computing and Artificial Intelligence Journal*. 2021. Vol.10. №3. P.227-240. DOI: 10.14201/ADCAIJ2021103227240.
- [18] **Eswara Rao K., Pydi B., Annan Naidu P., Prasann U.D., Anjaneyulu P.** Ensemble learning approach for effective software development effort estimation with future ranking. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*. 2023. №12(1). e31206. DOI: 10.14201/adcaij.31206.
- [19] **Ali Bou Nassif, Mohammad Azzeh, Ali Idri, Alain Abran.** Software development effort estimation using regression fuzzy models. *Computational Intelligence and Neuroscience*. 2019. 8367214. DOI: 10.1155/2019/8367214.
- [20] **Priya Varshini A.G., Anitha Kumari K., Vijayakumar Varadarajan.** Estimating software development efforts using a random forest-based stacked ensemble approach. *Electronics*. 2021. №10. 1195. DOI: 10.3390/electronics10101195.
- [21] **De Carvalho H.D.P., Fagundes R., Santos W.** Extreme learning machine applied to software development effort estimation. *IEEE Access*. 2021. Vol.9. P.92676-92687. DOI: 10.1109/ACCESS.2021.3091313.
- [22] **Sousa A.O. et al.** Applying machine learning to estimate the effort and duration of individual tasks in software projects. *IEEE Access*. 2023. Vol.11. P.89933-89946. DOI: 10.1109/ACCESS.2023.3307310.
- [23] **Jadhav A., Shandilya S.K., Izonin I., Gregus M.** Effective software effort estimation leveraging machine learning for digital transformation. *IEEE Access*. 2023. Vol.11. P.83523-83536. DOI: 10.1109/ACCESS.2023.3293432.
- [24] **Meharunnisa Saqlain M., Abid M., Awais M., Stević Ž.** Analysis of software effort estimation by machine learning techniques. *Ingénierie des Systèmes d'Information*. 2023. Vol.28. №6. P.1445-1457. DOI: 10.18280/isi.280602.
- [25] **Yasmin A., Haider W., Daud A., Banjar A.** Machine learning based software effort estimation using development-centric features for crowdsourcing platform. *Intelligent Data Analysis*. 2023. №28(1). P.299-329. DOI: 10.3233/IDA-237366.
- [26] **Mizanur Rahman, Partha Protim Roy, Mohammad Ali, Teresa Gonc, alves, Hasan Sarwar.** Software effort estimation using machine learning technique. *International Journal of Advanced Computer Science and Applications*. 2023. №14. DOI: 10.14569/IJACSA.2023.0140491.
- [27] **Akshay Jadhav, Shishir Kumar Shandilya.** Reliable machine learning models for estimating effective software development efforts: A comparative analysis. *Journal of Engineering Research*. 2023. Vol.11. Issue 4. P.362-376. DOI: 10.1016/j.jer.2023.100150.
- [28] **Priya Varshini A.G., Anitha Kumari K., Janani D., Soundariya S.** Comparative analysis of machine learning and deep learning algorithms for software effort estimation. *Journal of Physics: Conference Series*. 2021. Vol. 1767. №1. DOI: 10.1088/1742-6596/1767/1/012019.
- [29] **Nhung H.L.T.K., Van Hai V., Silhavy P., Prokopova Z., Silhavy R.** Incorporating statistical and machine learning techniques into the optimization of correction factors for software development effort estimation. *J Softw Evol Proc*. 2024. №6(5). e2611. DOI: 10.1002/smr.2611.

Сведения об авторах



Мурзагалеев Тагир Муратович, 1985 г. рождения. Окончил Томский политехнический университет (2008), к.х.н. (2012). Главный инженер проектов в ООО «РН-Проектирование Добыча» в области управления проектами и автоматизации процессов управления проектами. С 2024 года аспирант СПбГЭТУ «ЛЭТИ», по специальности «Искусственный интеллект и машинное обучение». ORCID: 0009-0004-1166-9487 tmurzagaleev@yandex.ru. ✉.

Жукова Наталия Александровна, 1983 г. рождения. Окончила СПбГЭТУ «ЛЭТИ» (2006), к.т.н. (2008). Ведущий научный сотрудник СПб ФИЦ РАН, доцент СПбГЭТУ «ЛЭТИ». Соавтор более 75 научных работ в области адаптивных измерительных процессов, интеллектуальных методов и алгоритмов анализа и оценки данных, технологий синтеза информации и знаний, обработки больших данных. ORCID: 0000-0001-5877-4461; Author ID (Scopus): 56406142300; Researcher ID (WoS): K-9143-2018. nazhukova@mail.ru.



Поступила в редакцию 28.02.2026, после рецензирования 11.05.2026. Принята к публикации 26.05.2026.



Ontology of machine learning methods in project management

© 2026, T.M. Murzagaleev^{1,2}✉, N.A. Zhukova³

¹ Saint Petersburg State Electrotechnical University "LETI" named after V.I. Ulyanov (Lenin) (ETU "LETI"), Saint Petersburg, Russia

² Limited Liability Company "RN-Proektirovanie Dobycha", Tomsk, Russia

³ Federal State Budgetary Institution of Science "Saint Petersburg Federal Research Center of the Russian Academy of Sciences", Saint Petersburg, Russia

Abstract

The application of machine learning methods in project management offers a number of potential advantages that can contribute to improving project management processes. Effective use of machine learning methods requires the identification of the most informative performance metrics while taking into account the diversity of available datasets, which in turn necessitates the structuring and systematization of project management elements. The aim of this study is to develop an ontology that provides the information necessary for the application of machine learning methods in project management, particularly for forecasting project effort and completion time. A review of machine learning methods, evaluation metrics, and datasets reported in scientific publications was conducted. Based on the results of this review, an ontology structure was developed and described for the application of machine learning methods in project management for effort estimation and project schedule forecasting. In addition, a scheme of user interaction with the ontology is presented. The results of the study may facilitate the process of identifying appropriate machine learning methods for project management tasks using new or publicly available datasets.

Keywords: ontology, project management, project schedule forecasting, project effort estimation, machine learning methods, dataset.

For citation: Murzagaleev T.M., Zhukova N.A. Ontology of machine learning methods in project management [In Russian]. *Ontology of designing*. 2026; 16(2): 300-313. DOI: 10.18287/2223-9537-2026-16-2-300-313.

Financial Support: This work was supported by the state budget, project No. FFZF-2025-0019.

Authors' contributions: Murzagaleev T.M. – source selection and analysis, ontology development; Zhukova N.A. – article structure development, scientific supervision, final conclusions.

Conflict of interest: The authors declare no conflict of interest.

List of figures and tables

Figure 1 – Representation of the class hierarchy in the Protégé editor

Figure 2 – Definition of "Object properties" in the Protégé editor using the example of "Has_metric"

Figure 3 – Individuals of the "Method_ML" class in the Protégé editor

Figure 4 – Data properties of the "Experiment" class in the Protégé editor

Figure 5 – Fragment of the graph describing classes, instances, and their relationships within the domain

Figure 6 – General graph representation of the ontology in Protégé

Figure 7 – Fragment of the ontology graph in Protégé

Figure 8 – SPARQL query in Protégé displaying methods and experiments

Figure 9 – SPARQL query in Protégé displaying experiments using the Random_forest method

Figure 10 – SPARQL query in Protégé displaying experiments in which $R2 > 0.90$ was used as the evaluation metric for machine learning model performance

Table 1 – Relationships between ontology classes through Object properties

Table 2 – Data properties of ontology class individuals

References

- [1] **Kalli Srinivasa Nageswara Prasad, Vijaya Saradhi MV.** Comprehensive project management framework using machine learning. *International Journal of Recent Technology and Engineering (IJRTE)*. 2019; 8(2S3). DOI: 10.35940/ijrte.B1256.0782S319.
- [2] **José Ignacio Santos, María Pereda, Virginia Ahedo, José Manuel Galán.** Explainable machine learning for project management control. *Computers & Industrial Engineering*. 2023; 180: 109261. DOI: 10.1016/j.cie.2023.109261.
- [3] **Forouzesh Nejad AA, Arabikhan F, Aheleroff S.** Optimizing project time and cost prediction using a hybrid XGBoost and simulated annealing algorithm. *Machines*. 2024; 12: 867. DOI: 10.3390/machines12120867.
- [4] **Dudarin PV, Tronin VG, Svyatov KV.** An approach to labor intensity evaluation in software development process based on neural networks [In Russian]. *Automation of control processes*. 2019; 3(57): 65–72. DOI: 10.35752/1991-2927-2019-3-57-65-72.
- [5] **Zakrani Abdelali, Hain Mustapha, Namir Abdelwahed.** Investigating the use of random forest in software effort estimation. *Procedia Computer Science*. 2019; 148: 343-352. DOI: 10.1016/j.procs.2019.01.042.
- [6] **Przemyslaw Pospieszny, Beata Czarnacka-Chrobot, Andrzej Kobylinski.** An effective approach for software project effort and duration estimation with machine learning algorithms. *Journal of Systems and Software*. 2018; 137: 184-196. DOI: 10.1016/j.jss.2017.11.066.
- [7] **Nassar AH, Elbisy AM.** A machine learning approach to predict time delays in marine construction projects. *Engineering, Technology & Applied Science Research*. 2024; 14(5): 16125–16134. DOI: 10.48084/etasr.8173.
- [8] **Napolskikh DL.** Developing a specialized ontology to represent the economic concept of a cluster [In Russian]. *Software products and systems*. 2024; 37(2): 155–163. DOI: 10.15827/0236-235X.142.155-163.
- [9] **Gruber TR.** A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993.
- [10] **Isaeva OS.** Building an ontology to systematize the characteristics of the Internet of Things network [In Russian]. *Ontology of Designing*. 2024; 14 (2): 243-255. DOI: 10.18287/2223-9537-2024-14-2-243-255.
- [11] **Gusenkov AM, Bukharaev NR, Biryaltsev EV.** Building subject domain ontology on the base of a logical data model [In Russian]. *Electronic libraries*. 2020; 23(3): 390-417. DOI: 10.26907/1562-5419-2020-23-3-390-417.
- [12] **Khalaf TZ, Çağlar H, Çağlar A, Hanoon AN.** Particle swarm optimization based approach for estimation of costs and duration of construction projects. *Civil Engineering Journal*. 2020; 6(2): 384–401. DOI: 10.28991/cej-2020-03091478.
- [13] **Soreti M. Liben, Demiss A. Belachew, Walied A. Elsaigh.** Comparing advanced and traditional machine learning algorithms for construction duration prediction: a case study of Addis Ababa's public sector. *Engineering Research Express*. 2024; 6(4): DOI: 10.1088/2631-8695/ad979f.
- [14] **Sklias G, Gkelios S, Dimitriou D, Sartzetaki M, Chatzichristofis SA.** Synergizing global and local strategies for dynamic project management: An Advanced Machine Learning-Enhanced Framework. *IEEE Access*. 2024; 12: 85955-85968. DOI: 10.1109/ACCESS.2024.3413890.
- [15] **Shen Zhang, Xuechun Li.** A comparative study of machine learning regression models for predicting construction duration. *Journal of Asian Architecture and Building Engineering*. 2023; 23(6). DOI: 10.1080/13467581.2023.2278887.
- [16] **Yeom DJ, Seo HM, Kim YJ, Cho CS, Kim Y.** Development of an approximate construction duration prediction model during the project planning phase for general office buildings. *Journal of Civil Engineering and Management*. 2018; 24(3): 238-253. DOI: 10.3846/jcem.2018.1646.
- [17] **Israr ur Rehmana, Zulfiqar Alib, Zahoor Jana.** An empirical analysis on software development efforts estimation in machine learning perspective. *Advances in Distributed Computing and Artificial Intelligence Journal*. 2021; 10(3): 227-240. DOI: 10.14201/ADCAIJ2021103227240.
- [18] **Eswara Rao K, Pydi B, Annan Naidu P, Prasann UD, Anjaneyulu P.** Ensemble learning approach for effective software development effort estimation with future ranking. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*. 2023; 12(1): e31206. DOI: 10.14201/adcaij.31206.
- [19] **Ali Bou Nassif, Mohammad Azzeh, Ali Idri, Alain Abran.** Software development effort estimation using regression fuzzy models. *Computational Intelligence and Neuroscience*. 2019; 8367214. DOI: 10.1155/2019/8367214.
- [20] **Priya Varshini AG, Anitha Kumari K, Varadarajan Vijayakumar.** Estimating software development efforts using a random forest-based stacked ensemble approach. *Electronics*. 2021; 10: 1195. DOI: 10.3390/electronics10101195.
- [21] **De Carvalho HDP, Fagundes R, Santos W.** Extreme learning machine applied to software development effort estimation. *IEEE Access*. 2021; 9: 92676-92687. DOI: 10.1109/ACCESS.2021.3091313.
- [22] **Sousa A.O. et al.** Applying machine learning to estimate the effort and duration of individual tasks in software projects. *IEEE Access*. 2023; 11: 89933-89946. DOI: 10.1109/ACCESS.2023.3307310.

- [23] **Jadhav A, Shandilya SK, Izonin I, Gregus M.** Effective software effort estimation leveraging machine learning for digital transformation. *IEEE Access*. 2023; 11: 83523-83536. DOI: 10.1109/ACCESS.2023.3293432.
- [24] **Meharunnisa Saqlain M, Abid M, Awais M, Stević Ž.** Analysis of software effort estimation by machine learning techniques. *Ingénierie des Systèmes d'Information*. 2023; 28(6): 1445-1457. DOI: 10.18280/isi.280602.
- [25] **Yasmin A, Haider W, Daud A, Banjar A.** Machine learning based software effort estimation using development-centric features for crowdsourcing platform. *Intelligent Data Analysis*. 2023; 28(1): 299-329. DOI: 10.3233/IDA-237366.
- [26] **Mizanur Rahman, Partha Protim Roy, Mohammad Ali, Teresa Gonc, Hasan Sarwar.** Software effort estimation using machine learning technique. *International Journal of Advanced Computer Science and Applications*. 2023; 14. DOI: 10.14569/IJACSA.2023.0140491.
- [27] **Akshay Jadhav, Shishir Kumar Shandilya.** Reliable machine learning models for estimating effective software development efforts: A comparative analysis. *Journal of Engineering Research*. 2023; 11(4): 362-376. DOI: 10.1016/j.jer.2023.100150.
- [28] **Priya Varshini A.G., Anitha Kumari K, Janani D, Soundariya S.** Comparative analysis of machine learning and deep learning algorithms for software effort estimation. *Journal of Physics: Conference Series*. 2021; 1767(1). DOI: 10.1088/1742-6596/1767/1/012019.
- [29] **Nhung HLTK, Van Hai V, Silhavy P, Prokopova Z, Silhavy R.** Incorporating statistical and machine learning techniques into the optimization of correction factors for software development effort estimation. *J Softw Evol Proc*. 2024; 6(5): e2611. DOI: 10.1002/smr.2611.
-

About the authors

Tagir Muratovich Murzagaleev (b. 1985) graduated from Tomsk Polytechnic University in 2008; Candidate of Chemical Sciences (2012). Chief Project Engineer at RN-Project Development and Production LLC, specializing in project management and the automation of project management processes. Since 2024, he has been a PhD student at Saint Petersburg Electrotechnical University “LETI” in the field of Artificial Intelligence and Machine Learning. ORCID: 0009-0004-1166-9487. tmurzagaleev@yandex.ru. ✉.

Natalia Alexandrovna Zhukova (b. 1983) graduated from Saint Petersburg Electrotechnical University “LETI” in 2006; Candidate of Technical Sciences (2008). Leading Researcher at St. Petersburg Federal Research Center of the Russian Academy of Sciences and Associate Professor at Saint Petersburg Electrotechnical University “LETI”. Co-author of more than 75 scientific publications in the fields of adaptive measurement processes, intelligent methods and algorithms for data analysis and evaluation, information and knowledge synthesis technologies, and big data processing. ORCID: 0000-0001-5877-4461; Author ID (Scopus): 56406142300; Researcher ID (WoS): K-9143-2018. nazhukova@mail.ru.

Received February 28, 2026. Revised May 11, 2026. Accepted May 26, 2026.
