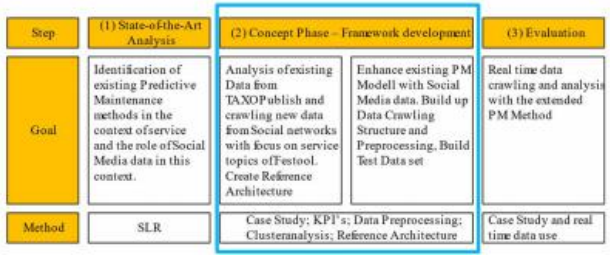
Reference Architecture framework for enhanced social media data analytics for Predictive Maintenance models	**Структура эталонной архитектуры для расширенного анализа данных в социальных сетях для моделей прогнозного обслуживания**
Jens Grambau Faculty of Print and Media Hochschule der Medien Stuttgart, Germany grambau@hdm-stuttgart.de	
Arno Hitzges Faculty of Print and Media Hochschule der Medien Stuttgart, Germany hitzges@hdm-stuttgart.de	
Boris Otto Faculty of Mechanical Engineering Technische Universität Dortmund Dortmund, Germany boris.otto@tu-dortmund.de	

Abstract— Social Media data contains a lot of hidden information which is currently rarely used in the manner of service topics on product level. However, a deep analysis of existing predictive maintenance models shows, that the combined use of social media data with already existing data from products or internal service data can improve existing and new analytical models for an enhancing predictive maintenance. Therefore, this framework paper describes an approach how to gather, process and analyze Social Media data related to products of a power tool producer. The defined processes are executed with the Azure Machine Learning Studio and are visualized with Power Bi. The main result of this paper is the Reference Architecture which combines several processes combine heterogenous data sources and enable the "First time to Incident" algorithm which helps companies to to increase the precision of Predictive Maintenance.	Аннотация. Данные социальных сетей содержат много скрытой информации, которая в настоящее время редко используется в качестве тем обслуживания на уровне продукта. Однако глубокий анализ существующих моделей прогнозного обслуживания показывает, что комбинированное использование данных социальных сетей с уже существующими данными о продуктах или внутренними данными обслуживания может улучшить существующие и новые аналитические модели для улучшения прогнозного обслуживания. Поэтому в этом базовом документе описан подход к сбору, обработке и анализу данных социальных сетей, связанных с продукцией производителя электроинструментов. Определенные процессы выполняются с помощью Azure Machine Learning Studio и визуализируются с помощью Power Bi. Основным результатом этой статьи является эталонная архитектура, которая объединяет несколько процессов, объединяет разнородные источники данных и включает алгоритм «Первый случай», который помогает компаниям повысить точность Прогнозного Обслуживания.
Keywords— Reference Architecture, Reference Model, IoT, Predictive Maintenance, Predictive Analytics, Predictive Model, Social Media, Data Mining, Big Data, Service, CRM, Framework, Visualization, Products and Services	Ключевые слова - эталонная архитектура, эталонная модель, IoT, прогнозное обслуживание, прогнозная аналитика, прогнозирующая модель, социальные сети, интеллектуальный анализ данных, большие данные, сервис, CRM, структура, визуализация, продукты и услуги
I. INTRODUCTION	I. ВСТУПЛЕНИЕ
In a world of digital transformation analytical solutions are becoming more and more important for companies. Grambau illustrates in the State-of-the-Art paper the latest approaches of Predictive Models in different industries with a focus on service-oriented topics. It can be seen that there are only a few approaches that link purely machine-generated data with human-made data from social networks, although this fusion could considerably increase the accuracy of forecasts [1]. However, this fusion cannot generally be applied to every product and service, because often there is no data available in the social networks or no connection of the data can be created. Grambau's analysis of existing Predictive Maintenance (PM) models clearly reveals the gap in research on a common	В мире цифровых преобразований аналитические решения становятся все более и более важными для компаний. В своей последней статье Грамбау иллюстрирует новейшие подходы к прогнозирующим моделям в различных отраслях с акцентом на сервис-ориентированные темы. Можно видеть, что существует только несколько подходов, которые связывают чисто машинные данные с созданными человеком данными из социальных сетей, хотя такое объединение может значительно повысить точность прогнозов [1]. Тем не менее, это объединение, как правило, не может применяться к каждому продукту и услуге, потому что часто нет данных, доступных в социальных сетях, или невозможно создать соединение данных. Анализ Грамбау существующих моделей

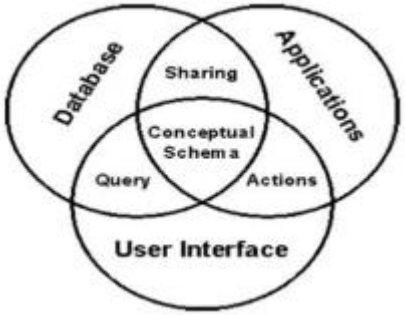
approach how to use Social Media data in the context of service for a PM method. Furthermore, it becomes clear, that for most of the Predictive Models in the B2B sector, the enrichment with Social Media data is difficult. In contrast, the B2C sector can profit a lot of Social Media data, as there is much more information available which is highly user specific and can be linked to specific products and services. Nevertheless, for B2B businesses it is recommended to search and use existing data in Social Media networks as the available information contains a lot of tacit knowledge about products or services. Moreover, the information is predominantly written by heavy users or subject matter experts.	Прогнозного Обслуживания (ПО) четко выявляет пробел в исследовании общего подхода к использованию данных социальных сетей в контексте обслуживания для метода ПО. Кроме того, становится ясно, что для большинства прогнозирующих моделей в секторе B2B обогащение данными социальных сетей затруднено. В отличие от этого, сектор B2C может извлечь выгоду из большого количества данных в социальных сетях, поскольку доступно гораздо больше информации, которая ориентирована на конкретного пользователя и может быть связана с конкретными продуктами и услугами. Тем не менее, для предприятий B2B рекомендуется искать и использовать существующие данные в социальных сетях, поскольку доступная информация содержит много скрытых знаний о продуктах или услугах. Кроме того, информация в основном написана активными пользователями или экспертами в данной области.
The new technical possibilities for the innovative and improved processing and use of data offer companies and their products as well as their services great potentials and synergies. The insights gained from the analysis of existing Predictive methods revealed the gap in existing research and justifies the need of the development of a new enhanced PM Model that additionally includes Social Media data as a new source of information. It is assumed that with the use of Social Media data PM Models can be more accurate and precise. This new data can provide a new view on products and services which then will lead to a better customer service. Not only the service maintenance of products can be enhanced but customer ideas also lead to even more innovative product development. Additionally, patterns can be derived from the analysis of the Social Media data. Additionally, this framework provides an implementation approach which can help existing PM Models to expand their analytical possibilities and features. Furthermore, companies can react earlier on customer needs and can improve their R&D for existing and new products. This framework helps companies to improve their overall data analytics approach as Social Media data is taken into consideration. This paper proposes a technical architecture based on a new developed reference architecture as well as an approach how to use and enhance existing data bases with Social Media	Новые технические возможности для инновационной и улучшенной обработки и использования данных сулят компаниям и их продуктам, а также их услугам большой потенциал и синергизм. Выводы, полученные в результате анализа существующих методов прогнозирования, выявили пробел в существующих исследованиях и обосновывают необходимость разработки новой улучшенной модели ПО, которая дополнительно включает данные социальных сетей в качестве нового источника информации. Предполагается, что с использованием данных социальных сетей модели ПО могут быть более точными и определенными. Эти новые данные могут дать новый взгляд на продукты и услуги, которые затем приведут к улучшению обслуживания клиентов. Не только сервисное обслуживание продуктов может быть улучшено, но идеи клиентов также приводят к еще более инновационной разработке продуктов. Кроме того, шаблоны могут быть получены из анализа данных социальных сетей. Кроме того, эта структура обеспечивает подход к реализации, который может помочь существующим моделям ПО расширить свои аналитические возможности и свойства. Так же, компании могут раньше реагировать на потребности клиентов и могут улучшить свои исследования и разработки для существующих и новых продуктов. Эта структура

data.	помогает компаниям улучшить общий подход к анализу данных с учетом данных социальных сетей. В этом документе предлагается техническая архитектура, основанная на новой разработанной эталонной архитектуре, а также подход к использованию и расширению существующих баз данных данными социальных сетей.
The structure of this this paper is organized as follows: In section II the objective, the research question and the research method are outlined. In the subsequent parts the reference model framework and structure are defined, and the PM framework is introduced. In section five the application of the reference architecture and the specific use case is described. In section six first results as well as hypotheses for the application of the use case are explained. In the conclusion main results are summarized and an outlook for future work is provided.	Структура этой статьи организована следующим образом: в разделе II изложены цель, вопрос исследования и метод исследования. В последующих частях определены структура и структура эталонной модели, а также представлена структура ПО. В пятом разделе описывается применение эталонной архитектуры и конкретный вариант использования. В шестом разделе объясняются первые результаты, а также гипотезы для применения варианта использования. В заключении обобщены основные результаты и дан прогноз на будущее.
II. OBJECTIVE, RESEARCH QUESTION AND METHOD	II. ЦЕЛЬ, ИССЛЕДОВАТЕЛЬСКИЙ ВОПРОС И МЕТОД
A. Objective and definition of research questions	A. Цель и определение вопросов исследования
As outlined above, preceding literature review of predictive methods revealed the lack of a general valid concept that describes the use of Social Media data in the manner of PM. Thus, this work aims in providing a description of a generalized PM Framework which shows the end-to-end process from gathering Social Media data, preprocessing and analyzing the data with Machine Learning methods, combining the data with machine-generated data and further visualizing and transferring the analyzed data. Such a generalized Framework could offer companies advantages in using Social Media data for their customer service as the additional access to Social Media data provides a new social view on services and products which in the end leads to a stronger relation and a better customer service.	Как указано выше, предыдущий обзор литературы по методам прогнозирования выявил отсутствие общей достоверной концепции, которая описывает использование данных социальных сетей в виде ПО. Таким образом, эта работа направлена на предоставление описания обобщенной структуры ПО, которая показывает сквозной процесс сбора данных в социальных сетях, предварительной обработки и анализа данных с помощью методов машинного обучения, объединения данных с данными, генерируемыми машиной, и дальнейшей визуализации и передачей проанализированных данных. Такая обобщенная структура может предложить компаниям преимущества в использовании данных социальных сетей для обслуживания своих клиентов, поскольку дополнительный доступ к данным социальных сетей обеспечивает новое социальное представление об услугах и продуктах, что в конечном итоге приводит к укреплению отношений и улучшению обслуживания клиентов.
In consequence, this work aims in answering the following research question	Следовательно, эта работа направлена на ответ на следующий вопрос

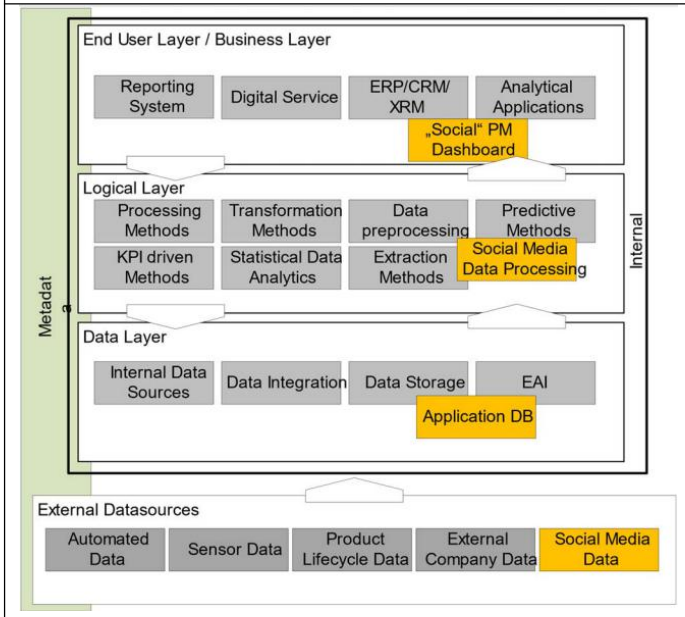
(RQ): • RQ 1: How should a PM Framework look like to use and include Social Media data and its hidden information for a PM application? This main research question can be specified into the following research sub questions (RSQ): • RSQ 1: How should the reference architecture of the enhanced PM Model be designed and how need the different layers to be adapted? • RSQ 2: How should the enhanced PM Framework process look like for being able to gather, preprocess and analyze Social Media data? • RSQ 3: How can the analyzed data be visualized and transferred to other systems that companies are able to use and benefit from the insights?	исследования (ВИ): • ВИ 1. Как должна выглядеть структура ПО для использования и включения данных социальных сетей и скрытой информации для приложения ПО? Этот основной вопрос исследования может быть указан в следующих подвопросах исследования (ПВИ): • ПВИ 1: как следует проектировать эталонную архитектуру улучшенной модели ПО и как необходимо адаптировать различные уровни? • ПВИ 2: Как должна выглядеть усовершенствованная структура ПО для возможности процесса сбора, предварительной обработки и анализа данных социальных сетей? • ПВИ 3: Каким образом анализируемые данные можно визуализировать и передавать в другие системы, которые компании могут использовать и извлекать пользу из полученных данных?
This paper refers to the second part of a dissertation, as shown in Fig. 1, the Concept Phase – Framework development and aims in developing of the PM framework based on the Reference Model for using Social Media data in this context.	Этот документ относится ко второй части диссертации, как показано на рис. 1: фаза разработки концепции структуры и направлен на разработку структуры ПО на основе Эталонной модели для использования данных социальных сетей в этом контексте.
 Fig. 1: Dissertation structured design, 2019.	Рис. 1: Структурный дизайн диссертации, 2019.
The outcome of the research is a new framework which can be applied to different industries and sectors for the enhanced use of Social Media data for predictable services. Thus, a transformative worldview can be enabled as there is no longer only a technical view, but also a social view on the data.	Результатом исследования является новая структура, которая может быть применена к различным отраслям и секторам для расширенного использования данных социальных сетей для предсказуемых услуг. Таким образом, с появлением социального взгляда на данные, а не только технического, может быть осуществлена трансформация

	мировоззрения.
The challenge of this research lies one the one hand in the nature of the data, as there are different sources with different heterogeneous formats available and on the other hand in the handling of the data to get a predictable outcome with the focus on the use cases.	Проблема этого исследования заключается, с одной стороны, в природе данных, поскольку существуют разные источники с разными гетерогенными форматами, а с другой стороны, в обработке данных, чтобы получить предсказуемый результат с акцентом на варианты использования.
The following sections of this paper outline the PM framework and its reference architecture, the PM Cockpit, and the PM Framework process. The main application of the PM Framework describes the case study and the generic process steps, the data collection, the data preprocessing, the data analyzation and the visualization of the results. In the result section the outcomes as well as hypotheses will be described which will be evaluated the continuative dissertation.	В следующих разделах этого документа описывается структура ПО и его эталонная архитектура, пульт управления ПО и процесс разработки структуры ПО. Основное применение структуры ПО описывает тематическое исследование и общие этапы процесса, сбор данных, предварительную обработку данных, анализ данных и визуализацию результатов. В разделе итогов будут описаны результаты, а также гипотезы, которые будут оцениваться в диссертации в дальнейшем.
B. Research Method	Б. Метод исследования
A common research strategy in the discipline of business informatics comprises the construction of a reference architecture. Schütte defines the construction process of a newreference model through a structured process divided into five phases [2].	Общая исследовательская стратегия в области бизнес-информатики включает построение эталонной архитектуры. Шютте определяет процесс построения новой эталонной модели посредством структурированного процесса, разделенного на пять фаз[2].
Within this work, a new reference model is created through following the five processes, that are described in detail below [3–5].	В рамках этой работы создается новая эталонная модель посредством следующих пяти процессов, которые подробно описаны ниже [3–5].
The first phase defines the basis and the common definition of the research question or the problem statement. It includes also a terminology which clearly identifies the attributes and objects which are used in the reference model. The research question defines the aim and the target of the new designed reference architecture.	Первый этап определяет основу и общее определение вопроса исследования или постановки задачи. Он также включает в себя терминологию, которая четко определяет атрибуты и объекты, которые используются в эталонной модели. Вопрос исследования определяет замысел и цель новой разработанной эталонной архитектуры.
A reference model framework defines a standardized model which is used to describe terms and elements of the framework on one basic language and reflects phase two. Schütte explains that the use of a so called “master-reference model” is useful for the universal structure and standardization.	Структура эталонной модели определяет стандартизированную модель, которая используется для описания терминов и элементов структуры на одном базовом языке и отражает второй этап. Шютте объясняет, что использование так называемой «эталонной модели» полезно для универсальной структуры и стандартизации.

Modelling languages are used in the third phase to formalize the framework which was defined in the second phase and create a so-called reference model structure. Therefore, detailed process and data modelling is done with usage of different languages like UML, EPK or ER-model. As the single languages are structured and consistent a syntactic and semantic correctness can be guaranteed. With a concrete case the processes and data models can be designed and be used for deducing of the new reference architecture.	Языки моделирования используются на третьем этапе для формализации структуры, которая была определена на втором этапе, и для создания так называемой структуры эталонной модели. Поэтому подробное моделирование процессов и данных выполняется с использованием различных языков, таких как UML, EPK или ER-модель. Поскольку отдельные языки структурированы и последовательны, могут быть гарантированы синтаксическая и семантическая корректность. В конкретном случае процессы и модели данных могут быть спроектированы и использованы для выведения новой эталонной архитектуры.
In phase four the case described and modelled in phase three will be generalized. The idea is to create from a casebased structure a deduction to a generalized reference model with all its processes. The generalized reference model can then be used and applied to different companies or cases. This enables to use benchmarking based on the reference model. This means basically that from an internal and external company point of view the new reference model is applicable	На четвертой фазе случай, описанный и смоделированный на третьей фазе, будет обобщен. Идея состоит в том, чтобы создать из основанной на кейсе структуры вывод к обобщенной эталонной модели со всеми ее процессами. Обобщенная эталонная модель может затем использоваться и применяться к различным компаниям или случаям. Это позволяет использовать опорные точки на основе эталонной модели. В основном это означает, что новая эталонная модель применима с внутренней и внешней точки зрения компании.
The application of the new generalized reference model reflects the fifth phase. It can be defined as the integration of the reference model in an existing environment of a company. The integration can vary from the implementation of single parts of processes or the aggregation of many reference models' parts to a lager reference model integration.	Пятый этап отражает применение новой обобщенной эталонной модели. Его можно определить, как интеграцию эталонной модели в существующую среду компании. Интеграция может варьироваться от реализации отдельных частей процессов или объединения частей многих эталонных моделей на лежаке эталонной модели интеграцию.
III. REFERENCE MODEL FRAMEWORK AND STRUCTURE	III. ОСНОВА И СТРУКТУРА ЭТАЛОННОЙ МОДЕЛИ
A. Construction of reference model framework	A. Построение структуры эталонной модели
The construction of a reference model is based on the consolidated view of knowledge. Schütte outlines the importance of using a defined language when constructing a reference model and additionally, recommends at a certain point of time a benchmark for the specific designed information system. Most of the designed reference models reflect process knowledge and thus, represent a good basic documentation that can be used for constructing the reference model framework. Moreover, the reference	Построение эталонной модели основано на сводном представлении знаний. Шютте подчеркивает важность использования определенного языка при построении эталонной модели и, кроме того, рекомендует в определенный момент времени ориентир для конкретной разработанной информационной системы. Большинство разработанных эталонных моделей отражают знания процесса и, таким образом, представляют собой хорошую базовую документацию, которая может использоваться

model can be developed out of a practical use case or out of research reflections [4].	для построения структуры эталонной модели. Кроме того, эталонная модель может быть разработана из практического варианта использования или из рефлексии исследования [4].
This research describes the way of using existing or new data sources for a new analysis of products and services in the context of customer service. As in many companies dedicated data bases are used for different systems like CRM, ERP, MD, etc. the ANSI-SPARC architecture is set as the base reference model framework. The basic three-level architecture provides the advantage to differentiate between the user view, the internal concept level and the internal database level. Furthermore, this three-level architecture can be simply interpreted as a schema shown in Fig. 2, where all layers can communicate with each other to process data in the best way and to extract the maximum achievable benefit from data and the conceptual processes. In the internal database level, the physical data structure is defined which enables an efficient access to data. The main part of the conceptual level shows how data is processed, what the interfaces are and where the data is saved.	Это исследование описывает способ использования существующих или новых источников данных для нового анализа продуктов и услуг в контексте обслуживания клиентов. Поскольку во многих компаниях выделенные базы данных используются для различных систем, таких как CRM, ERP, MD и т. д., в качестве основы базовой эталонной модели установлена ANSI-SPARC архитектура. Базовая трехуровневая архитектура обеспечивает преимущество, позволяющее различать представление пользователя, уровень внутренней концепции и уровень внутренней базы данных. Кроме того, эта трехуровневая архитектура может быть просто представлена как схема, показанная на рис.2, где все уровни могут взаимодействовать друг с другом, чтобы наилучшим образом обрабатывать данные и извлекать максимально достижимую выгоду из данных и концептуальных процессов. На уровне внутренней базы данных определяется физическая структура данных, которая обеспечивает эффективный доступ к данным. Основная часть концептуального уровня показывает, как обрабатываются данные, каковы интерфейсы и где хранятся данные.
 Fig. 2: ANSI-SPARC schema, [6]	Рис. 2: схема ANSI-SPARC, [6]
The Standard ASCI-SPARC architecture is transformed for this work and shows already a few new elements which will be developed for the specific construction of the new reference model structure. Fig. 3 provides an	Для этой работы стандартная архитектура ASCI-SPARC преобразуется и уже показывает несколько новых элементов, которые будут разработаны для конкретной конструкции новой структуры эталонной

architectural template of the reference architecture for the Social Media data enhanced PM Model. It is aligned to the standard ANSI-SPARC Architecture which was designed to combine and conclude all necessary parts of the framework for a common understanding [6] [7].	модели. На рис. 3 представлен архитектурный шаблон эталонной архитектуры для улучшенной модели ПО с данными из социальных сетей. Он соответствует стандартной архитектуре ANSI-SPARC, которая была разработана для объединения и заключения всех необходимых частей структуры для общего понимания [6] [7].
The reference architecture recommends the new view on the developed framework. There are several works available which mention key structures of architectures with different level types. Ong, et al. describes their data driven architecture in a five layer perspective as the Data source layer and Metadata layer can be aligned to the Data Layer and the External Data sources in this paper [8]. As enterprise systems or services are mandatory to use and work with the data also the connection between the different layers and processes are important [9]. The connection illustrated by Paz et al. are implemented in this architecture in a more streamlined way. The internal layers are anyhow connected in-between and not only in this three-layer view. A lot of architectures are basically defined by an internal part, where most of the data is processed, stored transferred and visualized, and by an external part, that mainly consists of external services to gather, transfer or store the data [10]. The reference architecture was enhanced in having an external data source as well as the Data Layer which contains all available data within a company. Wang describes the different architecture layers in three main layers: The consumer, the platform and the providers which is aligned to End User Layer, Logical and Data Layer and External Data sources Layer [11].	Эталонная архитектура рекомендует новый взгляд на разработанную платформу. Доступно несколько работ, в которых упоминаются ключевые структуры архитектур с разными типами уровней. Онг и др. описывает их управляемую данными архитектуру в пятиуровневой перспективе, поскольку уровень источника данных и уровень метаданных могут быть приведены в соответствие с уровнем данных и источниками внешних данных в этой статье [8]. Поскольку корпоративные системы или службы являются обязательными для использования и работы с данными, также важна связь между различными уровнями и процессами [9]. Соединение, проиллюстрированное Paz et al. реализованы в этой архитектуре более упорядоченным образом. Внутренние слои так или иначе связаны между собой, а не только в этом трехслойном виде. Многие архитектуры в основном определяются внутренней частью, где большая часть данных обрабатывается, хранится, передается и визуализируется, и внешней частью, которая в основном состоит из внешних служб для сбора, передачи или хранения данных [10]. Эталонная архитектура была усовершенствована благодаря наличию внешнего источника данных, а также уровня данных, который содержит все доступные данные внутри компании. Ван описывает различные уровни архитектуры в трех основных слоях: потребитель, платформа и провайдеры, которые согласованы с уровнем конечного пользователя, уровнем логического уровня и уровня данных и уровнем внешних источников данных [11].
As ANSI-SPARC was proposed in the midseventies it can be assumed that this architecture is outdated. However, in many modern RDBMS this architecture is still the basis for the construction and explanation of the database system landscape. The main benefit of the ANSI-SPARC architecture is the general point of view, that is used to describe an already	Поскольку ANSI-SPARC был предложен в середине семидесятых годов, можно предположить, что эта архитектура устарела. Однако во многих современных СУБД эта архитектура по-прежнему является основой для построения и объяснения ландшафта системы баз данных. Основным преимуществом архитектуры ANSI-SPARC является общая точка

existing system from a company in an easy way. Moreover, new data bases and its structures, new applications and its connection as well as new user interfaces can be designed and connected with the existing and new components [6].	зрения, которая используется для простого описания уже существующей системы компании. Более того, новые базы данных и их структуры, новые приложения и их соединения, а также новые пользовательские интерфейсы могут быть спроектированы и связаны с существующими и новыми компонентами [6].
By applying the ANSI-SPARC to this work, the basic reference model shown in Fig. 3 can be created in order to represent the structure of the cooperating company in the context of PM and the social web.	Применяя ANSI-SPARC к этой работе, можно создать базовую эталонную модель, показанную на рис. 3, чтобы представить структуру сотрудничающей компании в контексте PM и социальной сети.
 Fig. 3: Basic Reference Model derived from ANSI-SPARC, 2019.	Рис. 3: Базовая эталонная модель, полученная из ANSI-SPARC, 2019.
B. Construction of the reference model structure	B. Построение структуры эталонной модели
The reference model framework paired with the reference model structure is the basis for the reference architecture for the enhanced PM analytical approach. A variety of industry sector can be supported by using this new model and improve their data analytics. To design and describe the new structure one base language, the Unified modeling language (UML), is used. UML is one of the most common notations to design and describe software systems. Its base is a common notation and semantic to visualize,	Структура эталонной модели в сочетании со структурой эталонной модели является основой эталонной архитектуры для расширенного аналитического подхода PM. С помощью этой новой модели можно поддержать различные отрасли промышленности и улучшить их аналитику данных. Для разработки и описания новой структуры используется один базовый язык - унифицированный язык моделирования (UML). UML - одна из наиболее распространенных

construct and document models of business processes which should be available for objectoriented software development. There are 13 different types of diagrams to describe processes. For this work the type of activity diagrams is used, because procedures, processes and algorithms can be visualized and constructed [12–14].	систем обозначений для проектирования и описания программных систем. Его основой являются общие обозначения и семантика для визуализации, конструирования и документирования моделей бизнес-процессов, которые должны быть доступны для разработки программного обеспечения, не предназначенного для целей. Существует 13 различных типов диаграмм для описания процессов. Для этой работы используется тип диаграмм деятельности, поскольку процедуры, процессы и алгоритмы могут быть визуализированы и построены [12–14].
The use case is based on a power tool producer located in Germany. This company is chosen because of the huge amount of data which can be gathered in the internet or social web. Although the business model is highly B2B driven, end users communicate a lot in the internet. As the social web data related to this company is not yet fully gathered, analyzed and used in the purpose of PM, this represents a suitable case for this work. The reference model structure outlined in the following is based on available social data about this power tool producer. In this use case there are no existing internal processes for data analytics. Thus, the established reference model is designed from scratch and can be applied to any IT environment of other company.	Вариант использования основан на производителе электроинструмента, расположенном в Германии. Эта компания выбрана из-за огромного количества данных, которые могут быть собраны в Интернете или социальной сети. Хотя бизнес-модель основана на B2B, конечные пользователи много общаются в Интернете. Поскольку данные социальной сети, относящиеся к этой компании, еще не полностью собраны, проанализированы и использованы для целей ПО, это является подходящим случаем для этой работы. Структура эталонной модели, изложенная далее, основана на имеющихся социальных данных об этом производителе электроинструмента. В этом варианте использования не существует внутренних процессов для анализа данных. Таким образом, созданная эталонная модель разработана с нуля и может быть применена к любой ИТ-среде другой компании.
To build a new PM model different main processes are necessary which need to be implemented in the existing business processes of the company. The new designed processes can be divided into four main steps. In the beginning there is the data capturing process which is visualized in Fig. 4 and contains the description of the additional sources from the social web where specific product data will be gathered. A Social Listening tool (Microsoft Social Engagement) is used to gather specific data from open accessible social networks, blogs and newsfeeds. Furthermore, a web scrapper (WebHarvey) is used to gather product reviews from different platforms such as amazon, google shopping, CPO and further. The data can then be located at different places and must be transferred to one	Для построения новой модели PM необходимы различные основные процессы, которые необходимо внедрить в существующие бизнес-процессы компании. Новые разработанные процессы можно разделить на четыре основных этапа. Вначале идет процесс сбора данных, который представлен на рис. 4 и содержит описание дополнительных источников из социальной сети, где будут собираться данные о конкретных продуктах. Инструмент социального прослушивания (Microsoft Social Engagement) используется для сбора конкретных данных из открытых доступных социальных сетей, блогов и новостных лент. Кроме того, веб-скребок (WebHarvey) используется для сбора обзоров продуктов с разных платформ, таких как Amazon, Google

centralized local storage where the data can be processed. For this processing different interfaces are needed which are dependent on the used technologies. In this use case Azure Hub services are used to transfer the data from the Social Listening tool into an Azure SQL database and transferred to a local SQL database. The data from the web scrapper is saved directly in a local database.

Shopping, CPO и другие. Затем данные могут быть расположены в разных местах и должны быть переданы в одно централизованное локальное хранилище, где данные могут быть обработаны. Для этой обработки необходимы различные интерфейсы, которые зависят от используемых технологий. В этом случае службы Azure Hub используются для передачи данных из инструмента социального прослушивания в базу данных SQL Azure и переноса в локальную базу данных SQL. Данные из веб-хранилища сохраняются непосредственно в локальной базе данных.

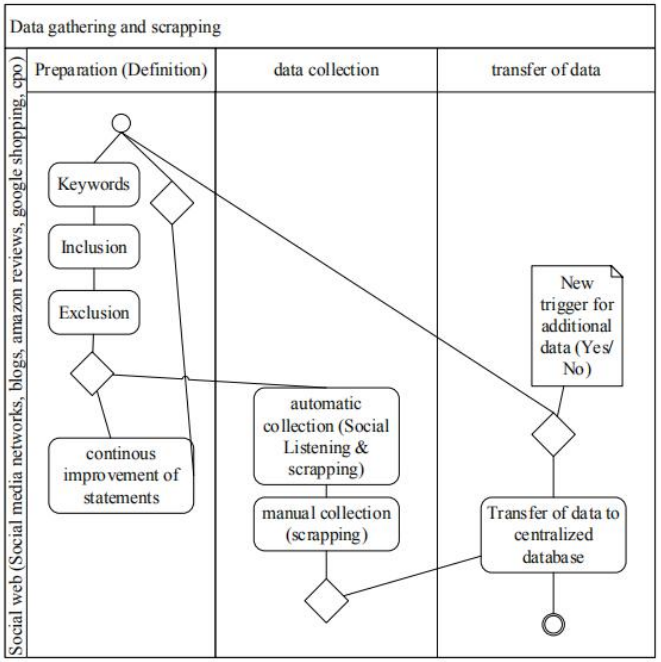


Fig. 4: UML data-capturing, 2019.

Рис. 4: UML-схема сбора данных, 2019 г.

In the second stage the gathered data must be preprocessed. There are several steps needed to process the data and make it available for the analytical part. First, there will be SQL statements executed which remove the Social Media data from unnecessary links, html tags and stop words. After this cleansing the SQL data will be transformed into a .csv file. The second main source of data are the product reviews from different shopping

На втором этапе собранные данные должны быть предварительно обработаны. Есть несколько шагов, необходимых для обработки данных и предоставления их для аналитической части. Во-первых, будут выполняться операторы SQL, которые удаляют из данных социальных сетей ненужные ссылки, HTML-теги и стоп-слова. После этой очистки данные SQL будут преобразованы в файл .csv. Вторым основным

or review websites. Additionally, there are several actions executed to clear out unnecessary metadata like stop words, html tags and emojis. A normalization of the text data will be done as the algorithms can work more efficiently in this lower-case text modules. In the last step the preprocessed data is exported to a .csv file. As Azure Machine Learning Studio (AML) is mainly used in this use case, the gathered data must be converted into the .csv file format for transferring and processing the data. Fig. 5 additionally shows webservice as a possible last step. A webservice can be used as part of an automated process that transfers data from one to next step.	источником данных являются обзоры продуктов с разных магазинов или обзорных веб-сайтов. Кроме того, есть несколько действий, выполняемых для удаления ненужных метаданных, таких как стоп-слова, HTML-теги и смайлики. Будет выполнена нормализация текстовых данных, поскольку алгоритмы могут работать более эффективно в этих строчных текстовых модулях. На последнем шаге предварительно обработанные данные экспортируются в файл .csv. Поскольку Azure Machine Learning Studio (AML) в основном используется в этом случае, собранные данные должны быть преобразованы в формат файла .csv для передачи и обработки данных. На рис. 5 дополнительно показан веб-сервис в качестве возможного последнего шага. Веб-сервис может использоваться как часть автоматизированного процесса, который передает данные от одного шага к следующему.

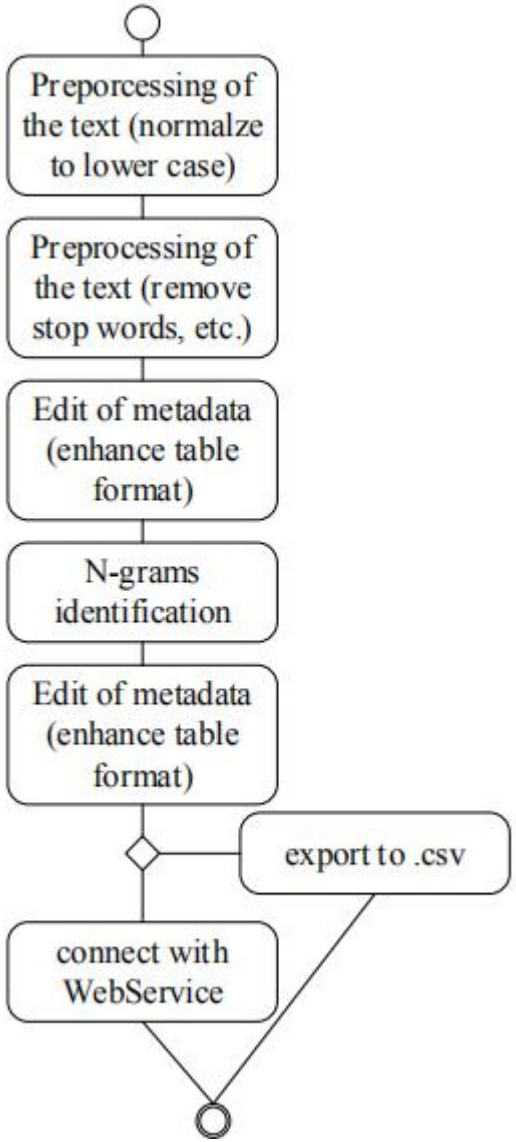
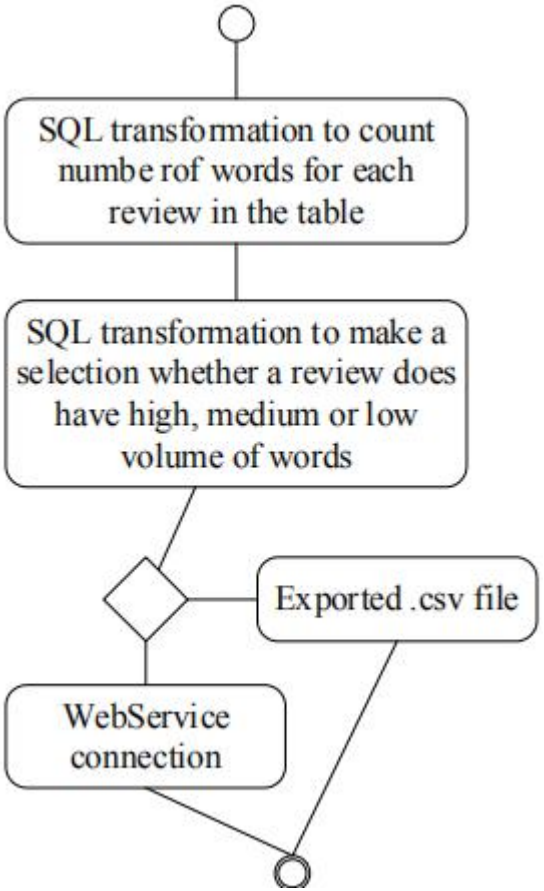


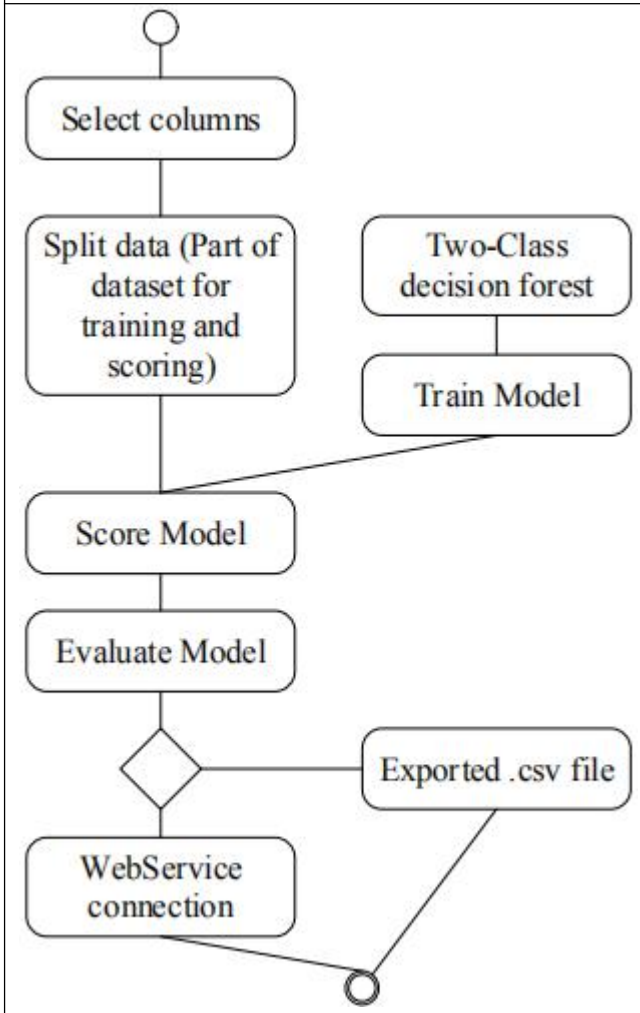
Fig. 5: UML data-preprocessing, 2019.

Рис. 5: UML-схема предварительной обработки данных, 2019.

The third process is a simple word count process that analyzes the size of the Social Media posts or the product reviews through counting the number of words. The number of words in a single data record is important for the analyzation in a later stage as there can be correlations deduced and

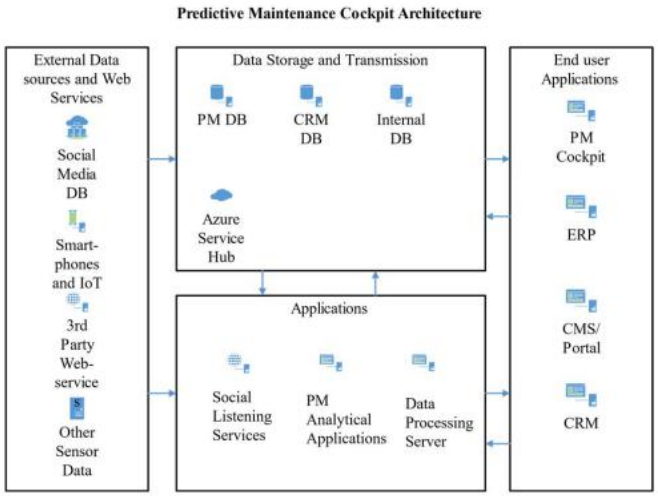
Третий процесс - это простой процесс подсчета слов, который анализирует размер сообщений в социальных сетях или обзорах продуктов путем подсчета количества слов. Количество слов в одной записи данных важно для анализа на более поздней стадии, так как

causalities be confirmed or declined. The analyzation of the number of words is executed by simplified SQL scripts as written in Fig. 6.	могут быть выведены корреляции и причинно-следственные связи подтверждены или отклонены. Анализ количества слов выполняется с помощью упрощенных сценариев SQL, как показано на рис. 6.
 Fig. 6: UML word-counts, 2019.	Рис. 6: UML-схема подсчета количества слов, 2019.
The application of the machine learning algorithms is based on the preprocessed data. In the social web many posts are not directly related to specific products. But more often they are related to a supplier, another vendor or the logistic company who delivers the product at the end. Therefore, it needs to be identified whether Social Media post, a review or any other data record is assigned to the product, the seller or other topics.	Применение алгоритмов машинного обучения основано на предварительно обработанных данных. В социальной сети многие посты не имеют прямого отношения к конкретным продуктам. Но чаще они связаны с поставщиком, другим поставщиком или логистической компанией, которая поставляет продукт в конце. Поэтому необходимо определить, назначена ли публикация в социальной сети, обзор или

The last process visualization in Fig. 7 shows the process steps needed for this identification. For this use case of a power tool producer a differentiation between the "product" and the "seller" is done. If there is no clear assignment possible, the post can be related to "both".	любая другая запись данных продукту, продавцу или другим темам. Последняя визуализация процесса на рис. 7 показывает шаги процесса, необходимые для этой идентификации. Для этого варианта использования производителя электроинструмента проводится различие между «продуктом» и «продавцом». Если нет четкого назначения, пост может быть связан с «обоими».
 Fig. 7: UML seller or product identification, 2019.	Рис. 7: UML-схема идентификации продавеца или продукта, 2019.
After the identification of the product or seller data records a sentiment	После идентификации данных о товаре или продавце будет проведен

analysis will be done. The sentiment analysis is based on a common sentiment database called "SENTIWORDNET" [15]. This database is one of the most common sentiment databases and open source based. The benefit of analyzing the sentiment with one common sentiment database is that one aligned base of the sentiment can be used. For the products of the power tool producer there are two main languages for the data records, English and German.	анализ настроений. Анализ настроений основан на общей базе данных настроений под названием «SENTIWORDNET» [15]. Эта база данных является одной из самых распространенных баз данных настроений и основана на открытых источниках. Преимущество анализа настроений с помощью одной общей базы данных настроений заключается в том, что можно использовать одну согласованную базу настроений. Для продуктов производителя электроинструментов существует два основных языка для записи данных: английский и немецкий.
For an enhanced PM approach new intelligent analytical method must be designed considering knowledge from existing methods. To include Social Media data in a PM model this data needs to be combined with IoT. It is important to identify and understand the cause-effect relationship from the Social Media and IoT data perspective. The PM Model will analyze continuously incoming product reviews as well as Social Media data like posts, tweets, statements and further posts that are related to errors, failures or damages of products. The words used for the analysis are defined in a new list which also includes words from a hierarchical reclamation list. It is important to update this list as soon as there is a new word used to describe certain circumstances of defects. The model analyzes data on a product level and determines how much and which posts are written in the web. If the number of posts will increase or decrease significantly, a notification will be sent immediately to the customer service team. Furthermore, the second PM process which is running is continuously analyzing the IoT data and compares those swings of the Social Media data with the ones from the mechanical data. The combination of those two data sources is then used for the visualization and the predictive approach of forecasting different cases like the needed service hours, the needed parts and the needed hyper care phase if a new product is released.	Для расширенного подхода PM должен быть разработан новый интеллектуальный аналитический метод с учетом знаний существующих методов. Чтобы включить данные социальных сетей в модель PM, эти данные необходимо объединить с IoT. Важно определить и понять причинно-следственную связь с точки зрения данных социальных сетей и IoT. Модель PM будет постоянно анализировать поступающие обзоры продуктов, а также данные социальных сетей, такие как сообщения, твиты, заявления и другие сообщения, связанные с ошибками, сбоями или повреждениями продуктов. Слова, используемые для анализа, определены в новом списке, который также включает слова из иерархического списка восстановления. Важно обновить этот список, как только появится новое слово для описания определенных обстоятельств дефектов. Модель анализирует данные на уровне продукта и определяет, сколько и какие сообщения пишутся в Интернете. Если количество постов будет значительно увеличиваться или уменьшаться, уведомление будет немедленно отправлено в службу поддержки клиентов. Кроме того, второй процесс PM, который выполняется, постоянно анализирует данные IoT и сравнивает эти колебания данных социальных сетей с колебаниями механических данных. Комбинация этих двух источников данных затем используется для визуализации и прогнозного подхода к прогнозированию различных случаев, таких как необходимые часы обслуживания, необходимые детали и необходимая фаза гиперобслуживания, если будет выпущен новый продукт.
IV. PREDICTIVE MAINTENACE FRAMEWORK	IV. ПЛАТФОРМА ПРОГНОЗНОГО ОБСЛУЖИВАНИЯ

Grambau outlines in the State-of-the-Art paper that the use of Social Media data for PM approaches is quite rare [1]. Moreover, most of these approaches use only machinegenerated data or only Social Media data. Additionally, there is no common method, but each approach uses different technologies and analyzation models to gain new insights from their data. Overall it is clearly shown, the more data is available the better the analyzation result is.	Грамбау отмечает в статье о состоянии дел, что использование данных в социальных сетях для подходов к ТМ встречается довольно редко [1]. Более того, большинство из этих подходов используют только данные, сгенерированные машиной, или только данные социальных сетей. Кроме того, не существует общего метода, но каждый подход использует различные технологии и модели анализа, чтобы получить новую информацию из своих данных. В целом это ясно показано, чем больше данных доступно, тем лучше результат анализа.
In this section the generic PM Framework is described which basically is a new concept to describe how to use Social Media data in a PM service cases.	В этом разделе описывается общая платформа ПО, которая в основном является новой концепцией, описывающей, как использовать данные социальных сетей в случаях прогнозного обслуживания.
A. PM Cockpit	A. ПО Пульт
Most of the existing approaches and technologies are aiming for gathering as much data as possible from technical sources, like sensor data, and to present an overview of a certain topic. The approach within this paper is to combine different data sources especially Social Media data with technical data. For this a new so-called PM Cockpit architecture is needed. The architecture in Fig. 8 describes the reference architecture transformed in the framework approach which consist out of four basic parts. The external data sources, the internal data sources and databases, the application software and the end-user application.	Большинство существующих подходов и технологий нацелены на сбор как можно большего количества данных из технических источников, таких как данные датчиков, и представление обзора по определенной теме. Подход в этой статье заключается в объединении различных источников данных, особенно данных социальных сетей, с техническими данными. Для этого нужна новая так называемая архитектура ПО Пульта. Архитектура на рис. 8 описывает эталонную архитектуру, преобразованную в базовый подход, который состоит из четырех основных частей. Внешние источники данных, внутренние источники данных и базы данных, прикладное программное обеспечение и приложение для конечного пользователя.

 Fig. 8: Technical architecture of the Predictive Maintenance Cockpit, 2019.	Рис. 8: Техническая архитектура пульта прогнозного обслуживания, 2019 г.
On the left side are the external data sources and web services. These are all publicly available information within and outside of Social Media networks and the world wide web itself. Moreover, sensor data from smartphones and IoT and other sensor data is considered. With different enterprise and application interfaces, Azure Service Hub technology and Social Listening tools the necessary data can be pulled and saved in the internal databases shown in the top box. Furthermore, the box shows the internal data sources: CRM and ERP as well as the custom internal databases, the PM database and specific databases for products and services of the company.	С левой стороны находятся внешние источники данных и веб-сервисы. Все это общедоступная информация внутри и вне сетей социальных сетей и самой всемирной паутины. Кроме того, рассматриваются данные датчиков от смартфонов, IoT и данные других датчиков. Благодаря различным интерфейсам предприятия и приложений, технологии Azure Service Hub и инструментам Social Listening необходимые данные можно извлекать и сохранять во внутренних базах данных, показанных в верхнем окне. Кроме того, в окне отображаются внутренние источники данных: CRM и ERP, а также настраиваемые внутренние базы данных, база данных ПО и конкретные базы данных для продуктов и услуг компании.
The second internal part is the application box. In here the data of the external and internal sources are processed. Social Listening applications crawl data from Social Media networks, the data processing application consolidate all the available data which is then transferred and processed within the PM application. In more detail the gathered data from the external and internal sources will be processed with specific software and algorithms and saved in a centralized database. The merged data will again be processed from machine learning software and algorithms, and the	Вторая внутренняя часть - это коробка для приложений. Здесь обрабатываются данные внешних и внутренних источников. Приложения Social Listening сканируют данные из сетей Social Media, приложение обработки данных объединяет все доступные данные, которые затем передаются и обрабатываются в приложении ПО. Более подробно собранные данные из внешних и внутренних источников будут обрабатываться с помощью специального программного обеспечения и алгоритмов и сохраняться в централизованной базе

results are provided to the end-user through CRM or the PM Cockpit.	данных. Объединенные данные будут снова обрабатываться из программного обеспечения и алгоритмов машинного обучения, а результаты предоставляются конечному пользователю через CRM или пульт ПО.
In the End-user Application box on the right all applications which can visualize the processed data are shown. Additionally, here the user can interact with the system and for example trigger a new request of a prediction.	В поле «Приложение конечного пользователя» справа отображаются все приложения, которые могут визуализировать обработанные данные. Кроме того, здесь пользователь может взаимодействовать с системой и, например, инициировать новый запрос прогнозирования.
Furthermore, it is highly important that compliance and data regulations as well as data laws within the European Union (EU) and the single countries where the tool should be used are considered. Many of the tools and applications are only available as a cloud solution and are located outside of the EU through which legal situation changes. This might be a risk for an implementation of a PM solution with the mentioned technologies in this paper. Therefore, an analysis has to be done to check the behavior and rights of the end user which work with the tool and which are providing information through Social Media or any other channel to a company. Moreover, there could be certain concerns and behaviors about the liberality of data of the people which differ from country to country.	Кроме того, очень важно учитывать нормативы соответствия и данных, а также законы о данных в Европейском союзе (ЕС) и в отдельных странах, где должен использоваться инструмент. Многие из инструментов и приложений доступны только в виде облачного решения и находятся за пределами ЕС, в результате чего меняется правовая ситуация. Это может быть риском для внедрения решения ПО с упомянутыми технологиями в этой статье. Поэтому необходимо провести анализ, чтобы проверить поведение и права конечного пользователя, который работает с инструментом и который предоставляет информацию через социальные сети или любой другой канал компании. Более того, могут быть определенные опасения и поведение в отношении либеральности данных о людях, которые отличаются от страны к стране.
For example, in Germany people are very skeptical about data transfers but not everybody reads the data privacy and the general terms and conditions or don’t think about using applications and websites. It’s a paradoxical situation which makes it necessary that the people are good informed what happen with their data, how it is processed and what is finally done with it.	Например, в Германии люди очень скептически относятся к передаче данных, но не все читают конфиденциальность данных и общие положения и условия или не думают об использовании приложений и веб-сайтов. Это парадоксальная ситуация, которая заставляет людей быть хорошо информированными о том, что происходит с их данными, как они обрабатываются и что в итоге делается с ними.
For the use of PM technology an efficient way of using available resources is mandatory. In most of the cases the resources from the cloud are used and rented for a certain time when the applications are running. For this it is necessary to plan the analytical processes to have to most appropriate costs in place.	Для использования технологии ПО эффективный способ использования доступных ресурсов является обязательным. В большинстве случаев ресурсы из облака используются и арендуются на определенное время при запуске приложений. Для этого необходимо спланировать аналитические процессы, чтобы обеспечить наиболее подходящие

	затраты.
Companies, their customers and end users benefit from the PM Cockpit in regards of products and services as they have more information available which results in an increasing quality of customer service activities. Service waiting times can be reduced and error sources are identified much faster. Furthermore, maintenance is better predictable, and the costs are reduced as unnecessary maintenance can be excluded in fact of eliminated error sources. End users also have the possibility to influence future product and service developments as their opinions and shared experiences are considered.	Компании, их клиенты и конечные пользователи получают выгоду от пульта ПО в отношении продуктов и услуг, поскольку они имеют больше доступной информации, что приводит к повышению качества обслуживания клиентов. Время ожидания обслуживания может быть уменьшено, и источники ошибок будут определены намного быстрее. Кроме того, обслуживание лучше предсказуемо, а затраты снижаются, поскольку ненужное обслуживание может быть исключено на самом деле из-за устраненных источников ошибок. Конечные пользователи также имеют возможность влиять на будущие разработки продуктов и услуг, учитывая их мнение и обмен опытом.
With this PM cockpit architecture following important facts are implemented: • Connection of heterogeneous data sources - Social Media data and the internal company data. • Identification of Social Media data's influence on predictions which only used machine generated data so far. Usage of new information to improve products and services through the experience of the customers. • Determination of improved predictions for service orientated topics of products or services. • Visualization of the filtered data in a so-called Service Cockpit or a CRM system.	С этой архитектурой пульта ПО, реализованы следующие важные факты: • Соединение разнородных источников данных - данные социальных сетей и внутренние данные компании. • Выявление влияния данных в социальных сетях на прогнозы, в которых до сих пор использовались только данные, сгенерированные машиной. Использование новой информации для улучшения продуктов и услуг через опыт клиентов. • Определение улучшенных прогнозов для сервис-ориентированных обсуждений продуктов или услуг. • Визуализация отфильтрованных данных в так называемой сервисоном пульте или системе CRM.
B. PM Framework Process	Б. Основа процесса ПО
The PM Cockpit is the base for the new PM Framework process which generically defines how the data is gathered, transferred via EAI/ API's to the internal databases, processed and visualized. If information and data is shared and created in real time in Social Media networks and analyzed afterwards, new insights and error sources might be detected. These errors could not be detected by analyzing only technical data.	Пульт управления ПО является базой для нового основного процесса ПО, который в общем определяет, как данные собираются, передаются через EAI / API во внутренние базы данных, обрабатываются и визуализируются. Если информация и данные передаются и создаются в реальном времени в социальных сетях и впоследствии анализируются, могут быть обнаружены новые идеи и источники ошибок. Эти ошибки

	не могут быть обнаружены путем анализа только технических данных.
The generic PM Framework consist out of four major process steps: • The gathering and filtering of the data from the different sources. • The processing of the data. Reducing redundancies and missing or invalid data. • Validation and training of the prediction algorithm based on specific parameters. This validation is an iterative process and thus, it can be reacted very fast to changing parameters and the model is trained. • In the last step all the validated data will be pushed and visualized to the PM database and be visualized in the cockpit or in a CRM system.	Базовая структура ПО состоит из четырех основных этапов процесса: • Сбор и фильтрация данных из разных источников. • Обработка данных. Сокращение избыточности и отсутствующих или недействительных данных. • Проверка и обучение алгоритму прогнозирования на основе конкретных параметров. Эта проверка является итеративным процессом и, таким образом, она может очень быстро реагировать на изменение параметров, и модель обучается. • На последнем шаге все проверенные данные будут переданы и визуализированы в базу данных ПО и визуализированы в пульте управления или в системе CRM.
The four major process steps are also reflecting the base architecture of the PM Cockpit. The gathering of data from external sources, the internal storage and processing of the gathered data and the visualization to the end-user.	Четыре основных этапа процесса также отражают базовую архитектуру пульта управления ПО. Сбор данных из внешних источников, внутреннее хранение и обработка собранных данных и визуализация для конечного пользователя.
The main process describes the overall picture of an enhanced process model for Social Media data in the manner of Predictive Models. This overview not shows the individual processes which are preprocessing and analyzing the data to be ready for the output.	Основной процесс описывает общую картину расширенной модели процесса для данных в социальных сетях в виде прогнозирующих моделей. Этот обзор не показывает отдельные процессы, которые предварительно обрабатывают и анализируют данные, чтобы быть готовыми к выводу.
C. Enrichment of existing predictive maintenance model	C. Обогащение существующей модели прогнозного обслуживания
The enrichment of an already existing PM method is useful as soon as there is an IoT integration set up. The idea behind this enhancement is a product related analyzation of data from the tool and data from the Social Media sources.	Обогащение уже существующего метода ПО полезно, как только будет настроена интеграция IoT. Идея этого усовершенствования заключается в анализе данных, полученных из продукта, и данных из социальных сетей.
To implement a combined solution the following components are needed. • IoT Infrastructure	Для реализации комбинированного решения необходимы следующие компоненты.

• Registered user and data collection of in field tools • In place analytics to preprocess data • Product group data hosting • API's • Agreement with user to process data	• IoT инфраструктура • Зарегистрированный пользователь и сбор данных в полевых инструментах • использовать аналитику для предварительной обработки данных • Хостинг данных группы продуктов • API • Соглашение с пользователем на обработку данных
If those major components are already in place or taken into consideration the data from Social Media can be gathered and implemented.	Если эти основные компоненты уже внедрены или приняты во внимание, данные из социальных сетей могут быть собраны и внедрены.
V. APPLICATION OF REFERENCE ARCHITECTURE	V. ПРИМЕНЕНИЕ СПРАВОЧНОЙ АРХИТЕКТУРЫ
In this section the experimental setup is described. The case study defines in which industry sector and in which specific company the reference model will be evaluated. The following table outline the participants, the considered data and the specific activities of the practical evaluation.	В этом разделе описана экспериментальная установка. Тематическое исследование определяет, в каком секторе промышленности и в какой конкретной компании будет оцениваться эталонная модель. В следующей таблице представлены участники, рассмотренные данные и конкретные мероприятия практической оценки.
The experimental setup is split into two phases and contains the necessary data sources which contain data over a defined timeframe beginning in January 2017 to April 2019. In phase one mainly the data preprocessing as well as a first analyzation with the tool Power BI is done. Power BI is used as it is a powerful tool which helps to analyze data in a visual way. There are already a lot of different examples how Power BI is used like for Sales Pipeline tracking in bigger companies or even in Universities to help students measure their performance [17].	Экспериментальная установка разделена на две фазы и содержит необходимые источники данных, которые содержат данные за определенный период времени, начиная с января 2017 года по апрель 2019 года. На первом этапе в основном выполняется предварительная обработка данных, а также первая аналитика с помощью инструмента Power BI. Power BI используется как мощный инструмент, который помогает визуально анализировать данные. Уже есть много разных примеров того, как Power BI используется, например, для отслеживания конвейера продаж в крупных компаниях или даже в университетах, чтобы помочь студентам измерить свою производительность [17].
TABLE 1: Case study definition Setting Power tool producer, a German company which produces power tools Participants • Social Media user • Online shop user • Company employees Included data • Internal	

Customer data • Social Media Posts • Shopping review data • Product data Activities • Gather data • Preprocess data • Analyze data • Visualize and transfer data	
The second phase has the focus on the analyzation of the data sets with machine learning algorithms. One of the tools is the Azure Machine Learning Studio which provides a huge amount of Machine Learning components [18]. The output and the results will then again be imported in Power BI for the better visualization.	Второй этап направлен на анализ наборов данных с помощью алгоритмов машинного обучения. Одним из инструментов является Студия машинного обучения Azure, которая предоставляет огромное количество компонентов машинного обучения [18]. Выходные данные и результаты будут снова импортированы в Power BI для лучшей визуализации.
VI. RESULTS AND EXPECTED RESULT HYPOTHESES FOR THE APPLICATION USE CASE	РЕЗУЛЬТАТЫ И ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ ГИПОТЕЗ ДЛЯ ПРИМЕНЕНИЯ ВАРИАНТОВ ИСПОЛЬЗОВАНИЯ
The findings taken from this work are formulated into hypotheses, correlations, causalities and KPI's. These aspects are evaluated in the final step of the dissertation. These statements have been derived from the first analysis with the anonymized data sets by using Power BI and the Azure Machine Learning Studio including the defined processes from section III. Furthermore, those statements help to understand the benefit of incorporating Social Media data in PM models.	Результаты, полученные в результате этой работы, сформулированы в виде гипотез, корреляций, причинно-следственных связей и ключевых показателей эффективности. Эти аспекты оцениваются на последнем этапе диссертации. Эти утверждения были получены в результате первого анализа анонимных наборов данных с использованием Power BI и Azure Machine Learning Studio, включая определенные процессы из раздела III. Кроме того, эти заявления помогают понять преимущества включения данных социальных сетей в модели ПО.
• With the analysis of the Social Media and review data over a certain timeframe (1 year, ½ year, 3 months) which is related to a product table the lifecycle of the products can be identified as well as patterns can be recognized which identify the occurrence of product failures. • A critical mass of posts can be identified which might indicate a product failure or defect and thus, the necessity of a service case. • The sentiment of a post is not necessarily in line with its rating. Thus, the sentiment analysis reveals additional information which might have been overlooked as the review is rated pretty good.	• С помощью анализа социальных сетей и анализа данных за определенный период времени (1 год, ½ года, 3 месяца), который связан с таблицей продуктов, можно определить жизненный цикл продуктов, а также шаблоны, которые идентифицируют возникновение сбоев продукта. • Может быть идентифицирована критическая масса постов, которая может указывать на неисправность или дефект продукта и, следовательно, на необходимость обслуживания. • Настроение поста не обязательно соответствует его рейтингу. Таким образом, анализ настроений показывает дополнительную информацию, которая могла быть упущена, так как обзор был оценен довольно хорошо.

In the analyzation part of the framework different correlations between the datasets themselves or between the data and the behavior of people in Social Media networks can be identified.	В аналитической части структуры могут быть выявлены различные корреляции между самими наборами данных или между данными и поведением людей в социальных сетях.
If in Social Media the number of posts for a specific product increases, the number of product related reviews will increase. • If the sentiment of Social Media posts is mainly negative, the amount of negative written reviews and received service cases will increase as well.	282/5000 Если в соцсетях количество постов по конкретному продукту увеличится, количество обзоров, связанных с продуктом, увеличится. • Если настроения в социальных сетях в основном отрицательные, количество отрицательных письменных отзывов и полученных случаев обслуживания также возрастет.
Moreover, based on the mentioned correlations and analysis of the data several causalities can be derived.	Кроме того, на основании упомянутых корреляций и анализа данных можно получить несколько причинно-следственных связей.
If the sentiment of Social Media posts is mainly positive, the number of written reviews and created service cases will decrease or stabilize on a low value. • If the number of Social Media posts with a positive sentiment increase, the number of positive written reviews will increase whereby the total number of reviews and the number of “negative” service requests will decrease. • The more product reviews and Social Media posts are generated the higher the potential for innovative product development and problem solutions is. • If the product reviews have a negative sentiment, the product rating will be low. • If the sentiment is negative, the number of words used in a single review will be higher than of a review with positive sentiment.	Если настроения в социальных сетях в основном положительные, количество письменных обзоров и созданных случаев обслуживания будет уменьшаться или стабилизироваться на низком уровне. • Если количество сообщений в социальных сетях с положительным настроением увеличится, количество положительных письменных отзывов увеличится, в результате чего общее количество отзывов и количество «отрицательных» запросов на обслуживание уменьшатся. • Чем больше обзоров продуктов и публикаций в социальных сетях генерируется, тем выше потенциал для разработки инновационных продуктов и решения проблем. • Если отзывы о товаре негативны, рейтинг товара будет низким. • Если настроение отрицательное, количество слов, использованных в одном обзоре, будет выше, чем в отзыве с положительным настроением.
The use of different technology tools enables the identification of different clusters inside the base data. With this possibility new clusters or even patterns in the base data as well as in the analyzed data can be detected.	Использование различных технологических инструментов позволяет идентифицировать различные кластеры внутри базы данных. С этой возможностью могут быть обнаружены новые кластеры или даже структуры в базовых данных, а также в проанализированных данных.

There will be a cluster where Influencer for power tools can be identified. • There will be a cluster which identifies the most responsive nations in the world.	Будет кластер, где можно определить влияние на электроинструменты. • Будет кластер, который идентифицирует самые отзывчивые страны в мире.
From a business perspective it is important to measure the analyzed data as well as interpret several conditions like "sentiment" as a KPI. The defined KPI's can be triggered from the business to measure the data which is available in the specific use case and additionally to identify the value of the data itself in the manner of a return on invest approach	С точки зрения бизнеса важно измерять анализируемые данные, а также интерпретировать некоторые условия, такие как «настроения», как KPI. Определенные KPI могут быть запущены из бизнеса для измерения данных, которые доступны в конкретном варианте использования, и дополнительно для определения ценности самих данных в виде подхода возврата инвестиций.
PR-Ratio: Number of posts in relation to the number of reviews. • PRSent-Ratio: The overall sentiment of Social Media posts in relation to overall sentiment of reviews. • PoRe-Div: The number of the Social Media post divided with the reviews of the specific product. The higher this value is, the higher is the potential that something happened with the product. • TTFI: Time to first service incident. Relation of IoT data to the amount over time of Social Media and review data.	PR-Ratio: количество постов по отношению к количеству отзывов. • PRSent-Ratio: общее настроение постов в социальных сетях по отношению к общему настроению отзывов. • PoRe-Div: количество постов в социальных сетях, разделенное на отзывы о конкретном продукте. Чем выше это значение, тем выше вероятность того, что что-то случилось с продуктом. • TTFI: время до первого сервисного инцидента. Отношение данных IoT к количеству сообщений в социальных сетях и количеству обзоров.
The mentioned KPI's are newly defined in the progress of this paper. The following two KPI's will be developed in the Azure Machine Learning Studio.	Упомянутые KPI вновь определены в ходе работы над этим документом. Следующие два ключевых показателя эффективности будут разработаны в Студии машинного обучения Azure.
• Rating of Keywords from different data sources like Twitter, Facebook, Amazon reviews, Google shopping reviews. • Time to first incident. Included in this mathematical calculation are the date of purchase, social media data, average value of the sentiment, number of service cases and average product review rating	• Рейтинг ключевых слов из разных источников данных, таких как Twitter, Facebook, обзоры Amazon, обзоры покупок Google. • Время до первого инцидента. В этот математический расчет включены дата покупки, данные социальных сетей, среднее значение настроения, количество случаев обслуживания и средний рейтинг отзывов о товаре.
VII. CONCLUSION	VII. ВЫВОД
The paper illustrates a framework to gather, process and analyze Social Media data in the manner of service-related topics. In this work the	В документе иллюстрируется структура для сбора, обработки и анализа данных социальных сетей в форме связанных с услугами обсуждений. В

established framework was applied to sample and anonymized product data from a power tool producer. However, the data was limited, significant insights could be gathered, and correlations and causalities be derived.	этой работе установленная структура была применена к образцу и анонимным данным о продукте от производителя электроинструмента. Несмотря на то что данные были ограничены, можно было собрать важные сведения и получить корреляции и причинно-следственные связи.
To verify the defined processes in this framework paper and evaluate the findings of section VI, the subsequent research will use the processes in a real-world use case with a power tool producer. New data sources as sensor data and further mechanical data will be integrated and combined with processes from this framework of the enhanced PM model.	Чтобы проверить определенные процессы в этом базовом документе и оценить выводы раздела VI, последующее исследование будет использовать процессы в реальном варианте использования с производителем электроинструмента. Новые источники данных, такие как данные датчиков и дополнительные механические данные, будут интегрированы и объединены с процессами из этой структуры расширенной модели ПО.
Future research will incorporate following components: • Cooperation with a power tool producer to define further use cases for the beneficial application of the enhanced PM model. • Sentiment analysis on a defined data base. • Evaluation and automation of the framework processes. • Validation of the Reference Architecture and the derived correlations and causalities. • Deriving prediction for specific products. Thus, continuous research will take the output of this work as a basis in order to establish a final and verified Predictive Maintenance model that incorporates Social Media data and aims in enhancement of service processes.	Будущие исследования будут включать следующие компоненты: • Сотрудничество с производителем электроинструмента для определения дальнейших вариантов использования для выгодного применения усовершенствованной модели ПО. • Анализ настроений на определенной базе данных. • Оценка и автоматизация базовых процессов. • Проверка эталонной архитектуры и полученных корреляций и причинно-следственных связей. • Получение прогноза для конкретных продуктов. Таким образом, дальнейшее исследование будет использовать результаты этой работы в качестве основы для создания окончательной и проверенной модели прогнозирующего обслуживания, которая включает в себя данные социальных сетей и направлена на улучшение процессов обслуживания.
REFERENCES	
[1] J. Grambau, A. Hitzges, and B. Otto, “Predictive Maintenance in the Context of Service - A State-of-theArt Analysis of Predictive Models and the Role of Social Media Data in this Context,” in ICEIS, 2018, pp. 223–230.	

[2] M. Rosemann and R. Schütte, “Multiperspektivische Referenzmodellierung: Scientific Figure on ResearchGate,” in Referenzmodellierung, J. Becker, M. Rosemann, and R. Schütte, Eds., Heidelberg: PhysicaVerlag HD, 1999, pp. 22–44.

[3] M. Maicher and H.-J. Scheruhn, Informationsmodellierung: Referenzmodelle und Werkzeuge. Wiesbaden: Deutscher Universitätsverlag, 1998.

[4] R. Schütte, Grundsätze ordnungsmäßiger Referenzmodellierung. Wiesbaden: Gabler Verlag, 1998.

[5] J. Vom Brocke, Referenzmodellierung: Gestaltung und Verteilung von Konstruktionsprozessen. Zugl.: Münster, Univ., Diss., 2002, 2nd ed. Berlin: Logos, 2015.

[6] G. Patel, “Study Of The ANSI/SPARC Architecture,” in International Journal of Modern Trends in Engineering and Research [Kurztitel fehlt!].

[7] ANSI-SPARC, 1975.

[8] I. Ong, P. Siew, and S. Wong, “A Five-Layered Business Intelligence Architecture,” CIBIMA, pp. 1– 11, 2011.

[9] J. F. de Paz et al., Eds., An Architecture for Proactive Maintenance in the Machinery Industry: Ambient Intelligence– Software and Applications – 8th International Symposium on Ambient Intelligence (ISAmI 2017): Springer International Publishing, 2017.

[10] H.-G. Kemper, W. Mehanna, and H. Baars, Business Intelligence - Grundlagen und praktische Anwendungen: Eine Einführung in die IT-basierte Managementunterstützung, 3rd ed. Wiesbaden: Vieweg + Teubner, 2010.

[11] L. Wang and X. V. Wang, Eds., Cloud-Based CyberPhysical Systems in Manufacturing. Cham: Springer International Publishing, 2018.

[12] G. Heinrich and K. Mairon, Objektorientierte Systemanalyse. München: Oldenbourg, 2008.

[13] A. Kemper and A. Eickler, Datenbanksysteme: Eine Einführung, 8th ed. München: Oldenbourg, 2011.

[14] K. C. Laudon, J. P. Laudon, and D. Schoder, Wirtschaftsinformatik: Eine Einführung, 3rd ed. Hallbergmoos: Pearson, 2016.

[15] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in LREC, 2010.

[16] J. Yang and A. M. Anwar, Eds., Social Media Analysis on Evaluating Organisational Performance a Railway Service Management Context. 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), 2016.

[17] J. Sluijter and M. Otten, "Business intelligence (BI) for personalized student dashboards," in Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, British Columbia, Canada: ACM, 2017, pp. 562–563.

[18] Microsoft Corporation, Microsoft Azure Machine Learning Studio. [Online] Available: https://azure.microsoft.com/de-de/services/machinelearning-studio/. Accessed on: Oct. 15 2018.