

**PAPUD****ITEA3 - 16037****Profiling and Analysis Platform Using Deep Learning****D6.1.6 Final Report**

Release : 1.0  
Status : Final  
Date : 13 November 2020  
Author : Cornel Crisan (Bull)  
Visibility : Public  
Pages : 51

**HISTORY OF CHANGES**

Version	Date	Changes	Author
V 0.1	27/10/2020	First draft	Cornel Crisan
V1.0	13/11/2020	Final version	Cornel Crisan

## Table of Contents

Table of Contents.....	2
1 Global Project synthesis.....	6
1.1 Project Identification .....	6
1.2 Project period identification .....	6
1.3 Partners.....	6
1.4 Project Objectives .....	7
2 Project Follow up .....	8
2.1 Project Planning .....	8
2.2 Work packages and deliverables .....	8
3 Work packages.....	11
3.1 WP1 – Models definition and development .....	11
3.2 WP2 – Models definition and development .....	12
3.3 WP3 – Product execution and test bed .....	12
3.4 WP4 – Validation.....	13
3.5 WP5 – Dissemination and exploitation.....	14
3.6 WP6 – Project management .....	15
4 Work done by the partners.....	17
4.1.1 4C & KU Leuven.....	17
4.1.2 BEIA .....	18
4.1.3 BULL .....	18
4.1.4 Ericsson .....	20
4.1.5 HI-Iberia .....	22
4.1.6 Institut Mines-Telecom (IMT) .....	23
4.1.7 Kocsistem .....	25
4.1.8 KU Leuven .....	26
4.1.9 Lille University.....	26
4.1.10 LORIA / University of Lorraine .....	27
4.1.11 Pertimm .....	29
4.1.12 ContentSide.....	30
4.1.13 Setur.....	31
4.1.14 Softeam.....	34

4.1.15	Turkgen .....	35
4.1.16	Turk Telekom .....	36
5	Partners efforts .....	38
6	Impact and Exploitation .....	39
6.1	Impact on the employment .....	39
6.1.1	4C .....	39
6.1.2	Beia.....	39
6.1.3	Bull .....	39
6.1.4	Ericsson .....	39
6.1.5	HI Iberia.....	40
6.1.6	Institut des Mines Telecom.....	40
6.1.7	Kocsistem .....	40
6.1.8	KU Leuven .....	41
6.1.9	Lille University.....	41
6.1.10	LORIA.....	41
6.1.11	Pertimm .....	41
6.1.12	ContentSide.....	41
6.1.13	Setur.....	41
6.1.14	Softeam .....	42
6.1.15	Turkgen .....	42
6.1.16	Turk Telekom .....	42
6.2	Market and exploitable results .....	42
6.2.1	Target markets and main drivers.....	42
6.2.2	Exploitable results.....	43
6.3	Individual exploitation plans.....	43
6.3.1	4C exploitation plan .....	43
6.3.2	Bull exploitation plan .....	43
6.3.3	Ericsson exploitation plan .....	44
6.3.4	HI Iberia exploitation plan.....	44
6.3.5	Pertimm exploitation plan .....	45
6.3.6	ContentSide.....	45
6.3.7	Turk Telekom exploitation plan .....	45
6.4	Open source approach.....	46

6.5	Communication and dissemination activities .....	46
6.5.1	Communication and scientific dissemination .....	47
6.5.1.1	Scientific publications .....	47
6.5.1.2	Publications accepted .....	47
6.5.1.3	Publications under review .....	48
6.5.1.4	Workshops and events.....	49
7	Conclusion.....	51

## List of Tables

Project Partners 1 .....	6
Work packages schedule 2.....	8
Deliverables Table 3.....	10

# 1 Global Project synthesis

## 1.1 Project Identification

Program: ITEA3 Call3

Project Acronym: PAPUD (ITEA 16037)

Start date: 15/12/2017

End date: 15/12/2020

## 1.2 Project period identification

This report covers the whole project, all the activities during the 3 years of this project.

## 1.3 Partners

Partner	Country	Type of Partner
Bull / ATOS	France	ind
4C	Belgium	SME
Hi-Iberia	Spain	SME
Beia	Romania	SME
Setur	Turkey	ind
Koc Sistem	Turkey	ind
Ericsson	Turkey	ind
KU Leuven University	Belgium	University
Press' Innov	France	SME
PERTIMM	France	SME
Softeam	France	ind
Mines Telecom	France	rsi
LORIA / University of Lorraine	France	University
Lille 1 University	France	University
Türk Telekom	Turkey	Ind
Turkgen	Turkey	SME

Project Partners 1

## 1.4 Project Objectives

As stated in the FPP (Full Project Proposal):

The world is experiencing a “Data Deluge” due to radical growth of data stemming from millions of heterogeneous, autonomous sources including social media, Internet of Things, the Web of Things, and various data-intensive industries. The data, regardless which media has been sourced from, have critical importance to any industry as well as for the quality of life. The “Data Deluge” enables Big Data technologies and Deep Learning by paving the way to do highly efficient analytics such as sentiment analytics, trend analysis, radicalization detection from text or various other media.

The scope of the PAPUD project (Profiling and Analysis Platform using Deep Learning) is to build models for data analytics in specific domains that are executed on a proposed set of technologies that fit best to the data provided. Indeed, before applying any kind of algorithm such as text summarization, translation or semantic analysis, one needs to build and learn the representative models for the data provided. The process of building representative models must deal with several complexity dimensions such as language, structure, semantic, etc.

While in lab settings, data characteristics are mastered, currently in the wild, businesses deal with heterogeneous data sets which cannot be dealt efficiently with current technologies. Today's necessity does not uniquely reside in the usage of independent data analytics processes; it is the utilization of combined technologies in which data analytics are executed to make sense out of the data. The business relevance of this project is guaranteed by the ability of realizing domain-specific analysis of structured and unstructured data. Indeed, we consider the applied industry requirements (such as high volatility, high volumes, dynamic data) and we built solutions that match to an optimal extent specificity of particular domains.

1. e-Commerce (including Social Media)
2. Call centers operations
3. Human Resources
4. e-Government and e-Services
5. Prescriptive Maintenance

## 2 Project Follow up

### 2.1 Project Planning

The PAPUD project is organized in 6 work packages as follow:

- WP1 Scope and requirements
- WP2 Models definition and development
- WP3 Product execution and test bed
- WP4 Validation and realization
- WP5 Commercialization and dissemination
- WP6 Project Management

Here below is the table of work packages involved during the project lifetime. Only the work package 1 “Scope and requirements”, the work package 2 “Models definition and development” and the work package 6 “Project management” are concerned by the first project semester.

	Q1 2018	Q2 2018	Q3 2018	Q4 2018	Q1 2019	Q2 2019	Q2 2019	Q2 2019	Q2 2019	Q2 2020	Q2 2020	Q2 2020
WP1												
WP2												
WP3												
WP4												
WP5												
WP6												

Work packages schedule 2

### 2.2 Work packages and deliverables

32 deliverables were provided including the 2 revisions of the state of the Art , 1 more than the number planned during the project lifetime. They are listed in the table below:



ID	TYPE	Name	Access	Due Month
D1.1.	Document	Market Analysis	Public	M6
D1.2	Document	Use cases definition	Confidential	M6
D1.3	Document	State of the Art	Public	M6
D1.3 V1.1	Document	State of the Art - Updated version	Public	M18
D1.3 V1.2	Document	State of the Art – Final version	Public	M30
D2.1	Document, Dataset	Data preparation	Confidential	M8
D2.2	Document, Software	Algorithm design	Confidential	M18
D2.3	Document, Software	Web services	Confidential	M30
D3.1.1	Hardware, software and Document	Deep Learning platform	Confidential	M9
D3.1.2	Hardware, software and Document	Deep Learning platform – Update	Confidential	M18
D3.1.3	Hardware, software and Document	Deep Learning platform – Final version	Confidential	M33
D3.1.4	Document	Hardware and software environments for big data processing and calculation frameworks	Confidential	M24
D3.2.1	Document	Use case installation and setup BAREM	Confidential	M16
D3.2.3	Document	Use case installation and setup e-commerce	Confidential	M16
D3.2.4	Document	Use case installation and setup – Call centers	Confidential	M16
D3.2.5	Document	Use case installation and setup – Human resources	Confidential	M16
D3.2.7	Document	Use case installation and setup - Log Analysis	Confidential	M16
D3.3	Document	Optimization and configuration change on the platform	Confidential	M33
D4.1.2	Document	Demonstrator and final report	Public	M34
D4.2.1	Document and software	Design and software implementation	Confidential	M18
D4.2.2	Document and software	Design and software implementation	Confidential	M36
D4.4.1	Document	Validation results for the use cases		M36

D5.1.1	Document	e-Citiz Product new feature sheet	Public	M36
D5.1.2	Document	Exploitation of DL models on Bull ATOS clusters	confidential	M36
D5.2.1	Document	CoterClub_PAPUD	Public	M36
D5.2.2	Document	Dissemination report	Public	M36
D6.1.1	Document	Progress report	Confidential	M6
D6.1.2	Document	Progress report	Confidential	M12
D6.1.3	Document	Progress report	Confidential	M18
D6.1.4	Document	Progress report	Confidential	M24
D6.1.5	Document	Progress report	Confidential	M30
D6.1.6	Document	Final report	Public	M36

Deliverables Table 3

## 3 Work packages

### 3.1 WP1 – Models definition and development

WP1 deals with the embodiment of the use cases by defining the roadmap and the required technologies. For each use case, a review is performed to identify all constraints related to users or a given domain. This work package will define all technical and functional requirements needed for the specification and the development step. Business-level strategies are also taken into consideration; the innovation will be determining factor for use cases study.

WP1 is built around 3 tasks:

#### T1.1 Business Model and Market Analysis (leader: HI-IBERIA)

This task will consist of business/market research and analysis that includes evaluation of best practices in the recruitment market. At the end of the task, a detailed report of the current market situation will be provided as well as a business model for the initial period of the project. According to changes in the market during de project lifetime this model might be updated as needed.

#### T1.2 Use case definition, requirements analysis and work methods (leader: BULL-ATOS)

This task has determined which case studies to put in place, what their needs are, what problems are faced and what solutions will be proposed. From the analysis carried out, the solution requirements will be extracted, depicting the functionalities to be provided.

Cases exploring the following domains:

1. e-Commerce (including Social Media)
2. Communications
3. Human Resources
4. e-Government and e-Services
5. Prescriptive Maintenance

#### T1.3 State of the art (leader: IMT)

This task compiles the existing State of the Art into a document to synthesize the overall purpose of the project results and their application to solve the existing problems. The document contains (a) the summarized State of the Art beyond what is captured in the PAPUD proposal covering technologies like NLP, text analysis , HPC and in summary different technologies around big data processing and machine/deep learning technologies; (b) a PAPUD manifesto in which the proposed solution is outlined without resorting to technical details mainly employed in the five use cases of the project.

Three deliverables, one for each task, have been produced.

WP1 was spread between January and June 2018 (M1 to M6).

An update of the State of the Art document has been produced in June 2020 taking in account the evolution of the research and technics in Deep Learning since the beginning of the project.

## 3.2 WP2 – Models definition and development

This work package will deal with the development of new approaches for use cases achievement. Indeed, each use case will bring a set of constraints in terms of data types and algorithms. As part of WP2, a review of all existing approaches and benchmarks will be performed to satisfy use cases requirements. Based on a theoretical model and experimental training, data scientists and developers will develop Deep Learning based approaches corresponding to the defined use cases. An important interaction will be considered with WP3 since all new solutions will be built on top of Big Data technologies such as Spark, MapReduce and related technologies.

The Task 2.1 concerns data preparation. All datasets have been collected and prepared and shared between partners (logs for HPC prescriptive maintenance, BAREM use case data, ...). All provided dataset are been used for experimentation. The corresponding deliverable (D2.1) has been produced in the second semester of 2018.

The Task 2.2 concerns algorithm designs. Basically, partners have worked to find the best Deep Learning model that feet each use case requirements. This task started in May 2018 and finished in June 2019. The deliverable (D2.2) has been produced.

The Task 2.3 is about algorithm deployment using Big Data platforms and more specifically on the web services specification and deployment. This task started at the end of 2018 and the deliverable (D2.3) has been produced in July 2020.

An update of the specifications has been done by Lille University in august 2020.

## 3.3 WP3 – Product execution and test bed

This Work Package deals with the HPC Hardware Platform and use cases environment set up. This execution environment is mandatory for the project and has been built in this work package.

The task 3.1 is about providing the Big Data and HPC Platform. It has started in July 2018. During the meeting in Leuven (June 2018) , partners had told what kind of tools they expect to use to give some insight of WP3 platform providers to design correctly the platform and ensure at least these tools will work on the platform. The Platform is up and running since end of december 2018 (M12). This task is ended and the deliverable D3.1 has been delivered. The deliverable D3.1.2 is about updates in the platform and it was delivered in M18. A final version of the enhancements brought to the platform is summarized in the deliverable D3.1.3 (September 2020)

The task 3.2 is about software installation and setting up. It includes use cases installation (if needed) and Deep Learning tools installation. The UC leaders have worked on the document collaborating with DL experts to identify softwares to install in the platform. This task has started in M9 (September 2018) and the milestone was reached with the delivery of the five D3.2 deliverables in M16 explaining how the platform will interact with use cases. This task is over.

The task 3.3 is about Platform optimization and enhancement. This task is running since the Use case installation on the platform (M16), each Use case gathered the enhancements and optimizations to do on the platform. The aim of this task is to customize and configure the platform to have better results. The final deliverable D3.3 has been delivered in October 2020.

### 3.4 WP4 – Validation

The objective of this WP is the validation of PAPUD platform regarding the use cases defined. The different tasks included in this WP includes the validation of the platform by means of the use cases defined. The demonstrators of the different use cases will be carried out with the outlined of the scenarios and the different demonstration elements, finally interoperability test will be executed to check if all technologies included in PAPUD could really be integrated and provide added-value, also data models will be crucial for the interoperability validation. Finally, five use cases have been defined in order to demonstrate the project aims within different domains. The use cases envisaged are the following: use Case 1: e-commerce; use Case 2: Call centre; use Case 3: Recommendation system for Human Resources; use Case 4: Behaviour Analysis for Reverse Efficient Modelling (BAREM); use Case 5: HPC

This WP started on December 2018 and it includes 4 tasks for use case demonstrations, design and dashboard implementation, and use cases validation led by HIB, ATOS, BEIA respectively.

Concretely during the first semester 2019, we have launched T4.1 by deploying the first version of demonstrators for the different use cases and we have been collecting information about them in an excel file including details such as demo status, data sets, DL algorithms deployed and initial performance results, and KPIs definition. Also, we have started working on T4.2, the deliverable is oriented on the design of the services on the platform, the missing bricks to have complete applications taking advantage of the DL algorithms.

During the second semester of 2019, we have been working in the preparation and deployment of demonstrators for the different UC to be ready for the ITEA Review. Also, we have delivered deliverable D4.2.1 with information about the HPC platform available in Bull and we have launched deliverable related to platform validation under task 4.3.

At the beginning of 2020 the consortium has been working in the improvement of the demonstrators based on the ITEA Y2 Review Results with the aim of delivering 5 demonstrators under the T4.1 with all the functionalities envisaged along the project. In addition to this, the different PAPUD use cases have been deployed within the PAPUD Platform provided by Bull following the directives in T4.2. Regarding T4.3 the test cases deliverable has been launched for the platform validation reporting. Initial contributions by the different use case leaders have been collected with the aim of reporting results obtained after running the DL and ML algorithms.

During this last semester 2020 we have worked on the validation results for each use case (D4.4.1) and on the demonstrator final report (D4.1.2)

### 3.5 WP5 – Dissemination and exploitation

The dissemination, the exploitation of results and the standardization are considered in WP5. Communication on the project is made at international level. Results generated by the project are published in renowned international scientific conferences and journals in order to rapidly reach other experts in the field.

In all technical reports, publications, presentations, demonstrators, and software, the project's ITEA support is explicitly mentioned and acknowledged. All dissemination actions such as the aforementioned technical deliverables, as well as those concerning communication media (website, interviews, invitations) give all due credit to the ITEA as sponsor.

Exploitation efforts will be supported by defining a new value with a business models and performing market studies for the case studies. Standardization includes the standardization of the all components and methodologies

Tasks:

T5.1: Exploitation of PAPUD results by new products, new features of existing products, ...

T5.2: Dissemination and communication of the PAPUD results (meetings, press conferences...)

WP5 has started on October 2018. A project website has been prepared and is being enriched <https://papud.wp.telecom-sudparis.eu/>. After one year, the partners have contributed 2 journal, 10 conference (IMT, BEIA, Lille) and 1 workshop publications (ATOS).

At month 18, the partners have contributed to 17 publications (journal, conference, workshop) and 10 events. ITEA review evaluation comments regarding individual exploitation plans have been taken into consideration.

At month 24, the partners have contributed to 27 publications (journal, conference, workshop) and 14 events. ITEA review evaluation comments regarding individual exploitation plans have been taken into consideration. In future reports, publications, presentations, software and media communication (website, interviews, invitations) the project's ITEA support will be explicitly mentioned and acknowledged.

During the last year of the project, due to Covid-19, we have faced some challenges attending conferences and organizing face to face meetings. Instead of our quarterly meeting we have conducted a series of online meetings where each partner have demonstrated their use case. At the moment, in total, partners have contributed to 29 publications (journal, conference, workshop) and 14 events. As we near the end of the project, partners are working towards their individual exploitation plans that will be detailed below.

### 3.6 WP6 – Project management

This work package deals with the management issue and includes the coordination between other work packages. Communications with ITEA are included in this work package. The project manager has the responsibility to provide regular reports about the project status, planning meeting, and review preparation.

We have 16 partners from 5 different countries. The work packages are shared by many partners. Maintaining communications clarity and meeting the deadlines has been a challenge with this large number of partners where the tasks are shared by several parties.

The project Kickoff has been organized by Bull in Grenoble (January 2018), the first quarterly meeting in Lille (march 2018) thanks to the University of Lille and the second quarterly meeting in Leuven (june 2018) with the help of KU Leuven and 4C. At the beginning of the project the number of partners has been fluctuant due the funding possibilities of partners in different countries. The portuguese partners finally left the project with no hope to be funded. This has somehow delayed the precise project definition with the relevant use cases. The project has really been in line after 6 months in June 2018.

Bull has organized the face to face meetings in Madrid in october 2018, and the meeting of Paris in February 2019 with the help of local partners and the work and support of the work package leaders WP2 (ContentSide), WP3 (Softeam) and WP4 (HI Iberia).

The work of coordination has been done with the work package leaders, the support of the task leaders, the country coordinators and also all the partners with weekly audioconferences, a sharepoint for the project and mails.

Bull has organized the first year review in Istanbul the 19th of march 2019 the with the active support of Kocsistem and the work and support of the work packages leaders. We have also organized a project Meeting in Bucharest June 2019 at Beia's offices.

And second change request has been done to take in account the reorganization of some tasks in the project.

A qaterly project Meeting aw organized in Nancy in October 2019 at LORIA's place and the second review in Istanbul in January 2020 thanks to Kocsistem.

Then the COVID 19 changed the game with no more possibilities to have face to face meetings. We replaced them by online meetings although the exchanges between partners are not as deep as during face to face meetings. The project stayed online and we a have realized the goals decided at the beginning delivering the planned documents, realizing the PAPUD Deep learning platform for the DL

training and applications, publishing articles on the innovations developed thanks to the project and preparing future products in the Uses case fields.



## 4 Work done by the partners

### 4.1.1 4C & KU Leuven

In close collaboration with the KU Leuven, 4C has built a multilingual solution, called TellMi, to automatically extract insights from text-based customer interactions to improve customer service (UC2).

Customers contact companies more and more through text-based channels. The analysis of these texts is often a manual process which makes it very time-consuming and expensive. And therefore, it is not often conducted. Moreover, companies in Belgium face an extra challenge because of the multiple official languages spoken within a relatively small client base. To automate text analysis, it is important to have large data sets. By dividing the group of customers based on language, a lot of potential knowledge is lost.

The objective of this use case is to automate the extraction and collection of the insights from customer interactions in different languages such as call logs, emails, webforms, ... . The platform TellMi gives an overview of what customers are talking about and how they feel about it across different languages and in real time.

Two models have been developed and integrated into the platform:

- **Cross-lingual aspect-based sentiment model:** This model does not only predict how the customers are feeling, but also the topic that evoked that sentiment. A deep learning multilingual aspect-based sentiment model was developed using word representations across different languages as input. To build these word representations, the KU Leuven built a cross-lingual vector space using the Incremental Hub Space technique, which was published in Heyman et al. (2019).
- **Cross-lingual fine-grained topic model:** Classical topic models result in a very high level of understanding in one language of what customers are talking about. Our solution provides a deeper level of topic definitions, which is streamlined across different languages. To detect the fine-grained topics, multi-lingual keywords were used as input of a deep learning topic model.

To perform the deep learning text analyses and store the data in a fast, scalable and affordable manner, TellMi has been developed on the cloud-based environments Heroku and AWS. TellMi has a user-friendly interface which allows companies to train models using their own customer interactions. The trained models can then be applied to new incoming interactions through the web application itself or by integrating it in their existing work environment through an API call. The TellMi application also provides a dashboard giving an overview of the predictions and the performance of the model.

Next, KU Leuven and 4C are developing a deep learning solution which automatically suggests possible responses to an incoming question of a customer, no matter which language he or she speaks. This will also be integrated in TellMi.

Reference to paper Heyman et al. (2019):

Heyman, G., Verreet, B., Vulić, I., & Moens, M. (2019). Learning unsupervised multilingual word embeddings with incremental multilingual hubs. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1 1890-1902. <https://doi.org/10.17863/CAM.39779>

#### 4.1.2 BEIA

For Beia, participating in the PAPUD project, together with partners, was a real privilege, a challenge at the same time and a source of inspiration for future projects.

Our contributions covered various aspects of the project, in accordance with our concerns and expertise, among which we can mention:

- voice-based application intended to a Call Center using KALDI28 in which routines for automatic speaker recognition and emotion detection have been developed.
- Market analysis and business model for UC3 – Recommendation system for Human Resources. Within this use case, BEIA provided data from online social networks which has been used as input by technology providers.
- We also contributed to the development of the PAPUD platform by integrating technologies for Natural language processing (NLP), Big Data analysis and machine learning.
- We have helped develop technologies for natural language processing (NLP), Big Data analysis and machine learning. Regarding NLP, Beia continues to focus on developing tools for processing the Romanian language.
- In order to facilitate Big Data processing, Beia is working on an application for analyzing and extracting relevant information from various sources, such as CVs and social networks. Thus Beia will be able to provide information on the specific needs and requirements of the Romanian recruitment market. To maximize the market penetration rate during operation, the PAPUD platform needs slight adaptations for specific markets. Beia will adapt the solution mainly for the Romanian market in order to increase its local socio-economic impact.
- We was pleased to be host for PAPUD F2F meeting with partners between June 20-21 in Bucharest, Romania.
- Also we can mention the dissemination activities during the events held in Romania, on which occasion we identified a real interest for the applications resulting from the project.

In conclusion, we are ready to exploit the results of the PAPUD project, both in current and future projects.

#### 4.1.3 BULL

The contribution of Bull to the project comprises 3 main tasks:

- **The management of the project** since Bull is leader of the international consortium and reports for all the PAPUD consortium to the ITEA. This role is detailed in the Work package 6 actions and realizations during the project lifetime.
- The elaboration of the **PAPUD platform** which allows the partners to run their Deep Learning trainings and inference inside their Web services. This includes the hardware conception of the platform, the installation of this platform inside Bull facilities, the selection and installation of the required software for Deep Learning , data preparation, web services for all the uses cases installed on this PAPUD platform. This platform for both hardware and software installations and setting has been described in detail in the deliverables D3.1.1, D3.1.2, D3.1.3 and D3.3. A Bull hardware and software administration team was specially supporting the partners for their software installation and use of this platform. Another Bull team has provided support for the Fast Machine Learning software framework developed by Bull and installed on the PAPUD platform.
- Bull has a **Use Case on HPC prescriptive maintenance**. In this use case Bull is data provider (HPC data from one client), Bull elaborates a DL model and tooling in collaboration with the LORIA.

**First Year** - At the beginning of the project Bull provided the data for this use case and developed a first solution for a simple use case for starting and cleaning the data. Regarding the State of the Art, Bull provided a non-exhaustive list of different methods and algorithms used for the Predictive Maintenance in HPC and several results concerning the prediction of errors. Then Bull focused on the preparation and the exploration of the data to build a first version of a data parser. Our first results show that we can predict the exit of an application with a precision of 60% and a recall of 80%.

**Year 2**- Bull added an action after the overheating prediction and beginning a new work about the log analysis. We are able to predict 70% of the overheating events. We design two preventive actions: a frequency reduction action and a migration task. These actions can help us to reduce the overheating cost by 65% in our case. In the second part, we split the system logs by "workflow" and provide a new visualization tool based on LSTM. Then we developed the web interface of the project and completed the development of the algorithms. To develop this interface, we merged the two parts of our tool. In addition, two publications were written.

**Year 3** - Our activity focused on the integration of our tools and the development of the prototype, including the web interface for the demonstration. The integration includes unit tests, test on several different hardware and dataset to ensure reliability in an industrial context. The web interface is now fully functional and can be used to show the results. The development of the tool has been completed, and it is published in open source. The last semester is devoted to the tools finalization (Logflow in Open source), the demonstration preparation and publications

**The project outcomes for Bull are:**

- The Deep Learning platform conception and realization both on hardware and software sides. This platform allows very different Uses cases to run in a “common” Deep learning environment. The portability thanks to the containers has been specially studied. The security of the applications through the Web has been also reinforced.
- The Deep Learning models and tools that will be integrated to the ATOS CODEX AI suite.
- The Doctoral Thesis presented by Marc Platini “Machine learning applied to the analysis and prediction of HPC systems failures”
- The fruitful collaboration with the project’s partners that will foster future projects.

#### 4.1.4 Ericsson

In the beginning of the PAPUD project, which Ericsson evaluated from the telecom field; it was aimed to make the textual forms of the voice calls to call centers to be anonymized and to run different algorithms on these data.

With these results, it was aimed to measure the caller's intention and satisfaction, as well as to measure the performance of the people answering the call or to work out a use case to prevent erroneous records.

It was targeted to work on deep learning techniques with call center’s data for early detection of potential churn subscribers by processing the data obtained from the written records of operator customers at call centers.

Due to GDPR rules, Ericsson had has to change the use case.

With the changes of the use case, Ericsson aimed to make anonymized the textual forms of voice calls coming to open sources and run different algorithms on taken data. Studies on public data collection were carried out by using different data in Ericsson data and algorithmic data production.

In line with these results, it was aimed to measure the caller's intention and satisfaction, as well as to measure the performance of the people answering the call and to study a use case to prevent erroneous records.

In order to develop this scenario, Ericsson has determined the analysis of requirements and working methods with Turkgen. Ericsson supported the examination of open source analysis frameworks.

*The outline of the work to be done is as follows:*

1) *Intent Classification:* Ericsson contributed with Turkgen the process of determining the reason for the customer to start the dialogue.

2) *Customer satisfaction measurement:* Ericsson contributed with Turkgen on Sentiment analysis which performed on different parts of the dialogue to determine how the dialogue begins, progresses and ends for the customer.

3) Agent performance measurement: Measurement criteria:

- a. Response times to the customer
- b. word rate misspelled,
- c. change in mood values (customer satisfaction) at the beginning and end of the dialogue.

4) Entity name recognition (Name Entity Recognition): In order to get rid of unnecessary variations / details in the processing of dialogs, entities such as person, product, company, brand names, prices, dates will be marked and brought to a generic status.

5) Correcting misspellings: In order to correct word mistakes and get rid of unnecessary variations, misspellings are determined with a spelling checker and then classical methods (edit distance, statistical language models) and neural language models (ex: Bidirectional language model) Encoder Representations from Transformers-BERT) are estimated and corrected.

2) Customer satisfaction measurement: emotional state on different parts of the dialogue.

How the dialogue started and progressed in terms of the customer by performing a sentiment analysis and its ending is determined.

3) Agent performance measurement: Measurement criteria:

- a. Response times to the customer
- b. Misspelled word rate, c. Mood values at the beginning and end of dialogues exchange (customer satisfaction).

4) Entity name recognition (Name Entity Recognition): Unnecessary in the processing of dialogs.

person, product, company, brand names, prices, to get rid of variations / details,

Entities such as dates will be marked and credited are still returned.

5) Correct spelling: To get rid of word mistakes, unnecessary variations

To make corrections, misspellings can be identified with a spelling checker and then the classic methods (edit distance, statistical language models) and neural language models (neural language models).

model (ex: Bidirectional Encoder Representations from Transformers-BERT) are corrected.

Ericsson provided set up environment for deep learning and high-performance interpretation , VM was created on the Ericsson Azure platform to enable the cloud installation of the services developed for the use-case in question

Ericsson has designed on Functional Platform Architecture configuration and optimization of the execution environment; post-deployment performance testing and architecture optimization were initiated.

#### 4.1.5 HI-Iberia

The contribution of HI-Iberia has been mainly focused on the following tasks:

- Management tasks → Due to the withdrawal of some partners at the beginning of the project HI-Iberia assumed the leadership of *WP4 – Validation and Realization, Task 4.1. Use Cases and Task 1.1 – Market analysis* which implies the coordination of the partners contribution, the organization of remote meetings for the work progress checking and the organization of WP related sessions during the face to face meetings. Regarding the use cases, HI-Iberia has been coordinating and managing the UC3-Human Resources Tool development within the project consortium. In addition to this HI-Iberia as Spanish country coordinator has been working in the management tasks related to national procedures and change requests.
- Technology tasks → The technology tasks developed along the project has been mainly focused on the development of the Human Resources Tool. The work done has been executed along the three years of the project covering from the analysis of the market and the acquisition of requirements until the development of Deep Learning algorithms and tool interfaces for the interaction within the Human Resources Department. Concretely:
  - Continuous organization of internal meetings at HI-Iberia with the Human Resources department to define the platform functionalities and requirements from those who will ultimately be the real users of the platform and to show the project results and improvements. Analysis of current tools available in the HR Department.
  - Market analysis for the Human Resources use case including: elaboration of the value chain, analysis of tools available in the market mainly in the CV parsing and recommendation of candidate domains, description of initial business model planned for the HR tool.
  - Analysis of the databases available in HI-Iberia (CVs and Job Offers) for elaborating the dataset. Development of the project datasets to train the algorithms.
  - State of the art of NLP techniques particularly for sentiment analysis, opinions, text analytics and multimedia content, deep learning methods in the HR area. Research on data analysis algorithms related to human resources that can serve as the basis for the PAPUD project from a theoretical perspective.
  - Definition of a common format for CVs and job offers that include the most relevant professional information for a job position in order to standardize the information from different sources within the platform.
  - Elaboration of a proper DL model as a basis for the analysis of pairs (CV, Offer) and training of the data model and evaluation of the results obtained. Initial training of the

- data model and evaluation of the results obtained after processing using the algorithms to have base results.
- Design and Development of web interfaces for interaction with the recommender by human resources personnel.
  - Development of a recommender for the Human Resources use case that, given a job offer, recommends the candidates that best fit the job. Several iterations have been carried out improving the results obtained in each previous iteration.
  - Continuous improvement of the initial dataset and the deep learning algorithm initially defined based on certain weak points detected in the tests of the first version of the recommender.
  - Unitary and interoperability tests for the different modules developed and, as a consequence, detection of the strengths and weaknesses of each of them in order to provide feedback for future improvements. Validation of the functionalities of the tool internally at HI-Iberia with the available CVs.
  - Definition of the software and configuration requirements to deploy and execute the algorithms of the human resources use case on the platform provided by ATOS Bull for the project. Installation and deployment of a docker on the PAPUD platform for the Human Resources use case where the DL models developed for the use case will be executed.
  - Analysis and research on the information available in EUROSTAT and INE (Spanish Statistics Institute) to analyze trends in the IT market.
  - Development of a time series model for providing the current market situation related to salaries at some job positions related to the IT market. Development of the user interfaces to visualize the market trends in a graph.
  - Technical preparation of materials (prototypes integrating the different modules developed) for the different meetings including review meetings with ITEA board.

Thanks to the work performed through the PAPUD project, HI-Iberia has obtained a Human Resources Tool developed over a Deep Learning algorithm to be exploited within the Human Resources Departments of IT companies. Also, the collaborative work executed has brought HI-Iberia the opportunity of collaborating with some partners from other countries and so to extend its contacts network by creating synergies for future work and project collaborations

#### 4.1.6 Institut Mines-Telecom (IMT)

As an academic partner in PAPUD project, IMT started its work by contributing in:

**WP 1 – Scope and requirements:** IMT contributed to D1.1 in the market analysis of two use cases including UC1 (e-commerce) and UC4 (BAREM). We reviewed the deliverable D1.1 as one of the reviewers assigned to this deliverable. We also contributed to D1.2 in use case definition comprising UC1 and UC4 as technology and data providers. IMT took the lead of T1.3, provided the ToC of the SoTA for this deliverable, and contributed to provide the content of different sections in this deliverable

including big data and online social networks analytics, and the SoTA related to UC1 and UC4. We delivered the final version of this deliverable by July 3, 2018. We also provided content for the updated version of SoTA deliverable including natural language processing and word embeddings techniques by June 3, 2020.

**WP2 - Models definition and development:** IMT was the leader of task T2.1 and completed the dataset preparation with the help of partners. IMT also contributed in definition of the functional architecture and integration of different parts of the two use cases UC1 and UC4. IMT tried to collect data for UC4 from social media (e.g. Twitter). However, due to the lack of required data for UC4, during the rest of the project, IMT was not able to contribute in this use case anymore.

Regarding task T2.2, algorithm design, IMT designed a deep learning algorithm for recommender system. We took a dataset in 4GB in size from Setur and cleaned/prepared the data in the format of our algorithm input (transactions: sequences of products). We extracted product features embedding using a shallow neural network, according to word2vec approach in text mining area, called product2vec. The result was integrated with the algorithm from e-commerce partner in UC1 (Kocsistem) as one of the inputs of their algorithm. IMT also proposed a new deep learning model for predicting the popularity of products inspiring the language modeling approach. We developed a Recurrent Neural Network (RNN) model based on Gated recurrent units (GRUs) and achieved Mean Average Precision (MAP) value 80%. This module is integrating with recommendation engine module in UC1 to sort the recommended products according to their popularity; therefore, there will be more chance for users to buy products that are more popular. Furthermore, the popularity scores of all the products in Setur dataset was integrated with search engine module from Pertim, which resulted in an improvement in the performance of the search engine.

In addition, IMT contributed on D2.2 to provide the content of related work, implementation details, and evaluation results sections. We also reviewed the deliverable D2.2 as one of the reviewers assigned to this deliverable. Regarding T2.3, we provided our web service description along with recommender system and search engine descriptions in UC1, and reviewed the D2.3 report on web services as internal reviewer.

**WP3 - Product Execution and Test Bed:** IMT contributed on deliverable D3.2.3 by providing the content to describe deep learning libraries that are mandatory for the UC1 and how to install these libraries. As two modules from IMT were integrated with recommendation engine from Kocsistem, they are accessible through Nova platform according to the recommendation engine configuration in Nova platform.

IMT also reviewed the deliverable D.3.2.5 as one of the reviewers assigned to this deliverable.

**WP4 Validation and Realization:** IMT contributed on D4.1 by including current state of the result and target value of improvement during the project. After integrating the popularity module with recommendation engine, the final evaluation scores of the modules are reported.

**WP5 - Dissemination, Commercialization and Standardization:** As an academic partner, IMT contributed to T5.2 by disseminating the project technical outcomes to relevant venues. Since the



beginning of the project, we have published some conference/journal papers in IEEE ICC, Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), IEEE Global Communications Conference, and IEEE Transactions on Smart Grid.

IMT has provided a domain from Telecom-SudParis and designed and developed a website for the project, which is accessible through: <https://papud.wp.telecom-sudparis.eu>. It is including project context and description, partners' description, disseminations, publications, and events during the project lifetime. We are also keeping update the website with dissemination and publication from partners during the project.

#### 4.1.7 Kocsistem

The main contributions of KoçSistem (KS) during the project timeline was on recommendation engine development for e-commerce use case (UC1). Kocsistem took active part in all work packages.

**Problem definition:** KS has participated in all three deliverables of WP1, providing the state of the art on deep learning for recommendation engines and taking active part in planning and defining the e-commerce use.

**Data:** KS has conducted the task of data preparation, preprocessing and analysis of Setur duty free data set. They have also performed benchmarking calculations with the data.

**Deep Learning Environment:** KS has taken part in setting up a deep learning environment for the recommendation engine both on premises and on the Nova platform. KS has completed the design of specifications of the docker environment and has set it up on Nova platform for the use of recommendation engine. During the Project timeline, Nova platform was actively used for both model training and inference.

**Web Service:** KS has developed a web service on Nova platform for the recommendation engine that was integrated by Pertimm to their search engine. Different deep learning algorithms that were developed during the project are integrated in the web service. The web service uses modules from both KS and IMT algorithms.

**Deep Learning Algorithms:** KS has developed novel deep learning algorithms for e-commerce use case's recommendation system. KS has designed unsupervised and supervised architectures to create the product and basket embeddings for a recommender system for both online e-commerce site and duty-free shops. The main problem addressed in the use case was learning product and basket representation for non-personalized baskets without using user information. The neural networks that were designed for this purpose transform baskets and products more successfully than other traditional approaches. A deep learning-based supervised model based on a feed-forward neural network was developed to address the problem and the dataset. For comparison, three baseline algorithms based on randomization and co-occurrence, two unsupervised models based on product embeddings was developed. Experiments indicate that the proposed deep learning architecture shows better performance than baseline approaches in terms of many metrics. Our deep learning architecture outperformed the unsupervised models with 11.8 % accuracy improvements and 4.0 MRR metric improvements.

**Collaboration:** During the development of the recommendation engine, a strong international collaboration was conducted between KS (DL algorithms and web service), IMT (DL algorithms), Pertimm (search & recommendation engine site) and Setur (Data). A final demonstration on Nova platform is prepared by partners that uses all components mentioned above. An academic collaboration was also started in Turkey with Bilgi University. It has resulted in successful development of DL models, and further academic collaboration on different projects. A recent paper from the results of the recommendation engine study, titled 'A Non-Personalized Recommender System by Learning Product and Basket Representation', is recently submitted to Turk J Elec Eng & Comp Sci.

**Project Management:** KS was WP5 and T2.3 (of WP2) leader. KS followed up the timely completion of tasks and deliverables of WP5 and T2.3, also took part in internal review of deliverables, preparation of documents and reports and participated in internal meetings. KS was also country leader of the Turkish national consortium. They have established the necessary coordination for completion of tasks, national reviews, internal and international communication. KS hosted first two of the ITEA reviews on its premises.

#### 4.1.8 KU Leuven

See above the common report with 4C.

#### 4.1.9 Lille University

Lille University efforts were focused on video and multi-sensors dedicated deep learning algorithms for affect estimation with particular attention to the frustration detection in the BAREM use case.

**BAREM Use Case** Lille University had the responsibility of a pilot study and associated capturing tools for recording user interactions while using the bus reservation system provided by SofTeam. This pilot study is the first one in the literature mixing together video, audio, bio-sensors and navigation logs. Twenty multi-sensors recordings sessions were completed. Tools for sanitizing and annotating the data have been implemented. Annotated session log data have been provided to the consortium. GDPR issues were also dealt with in order to protect the collected data. In parallel, deep architectures have been explored for exploiting visual and sensor data and spot changes in user behaviour during the BAREM experiment. We notably explored deep learning architectures for anomaly detection in order to distinguish frustration versus normal behaviour on the data collected within the pilot, but also on an existing video-only dataset named DAISEE. All models (CNN, RNN) are available<sup>1</sup> and they were deployed successfully on the PAPUD HPC platform.

Beyond the direct contribution to the BAREM use case, Lille University contributed mainly to the development of innovative algorithms for dealing with challenges that are often met in real-life situations similar to those envisioned in the BAREM use case. We have mainly addressed: face registration in presence of movement, expression intensity variations, and facial occlusions.

**Face registration** Facial expression recognition often requires near-frontal views. However, these situations do not fairly represent real-world challenges, where the subject is free of its movement. This

---

<sup>1</sup>[https://gitlab.univ-lille.fr/fox/papud\\_barem\\_frustration](https://gitlab.univ-lille.fr/fox/papud_barem_frustration)

is reflected in the accuracy drop of facial expression methods obtained on recent databases. Two challenges (head pose variations and large displacements) in facial expression recognition were studied. Experiments are proposed in order to quantify the impact of free head movements using representative expression recognition approaches (LBP, LBP-TOP, HOOF). We propose an experimental protocol (SNaP-2DFe) that records, under controlled light, facial expressions with two cameras: one attached on the head and one placed in front of the subject. As in both cameras facial expressions are the same, differences in performances measured on each camera show the impact of head pose variations and large displacements on the underlying recognition approach.

**Expression intensity and facial movement** In order to cope with a wide range of emotion intensities we consider the facial movement as a first class solution. However, facial movements are often subject to noise due to the homogenous texture and aperture problem, it is difficult to separate consistent movement patterns from noise. Direction and magnitude statistical profiles are jointly analysed in order to filter out noise. Various intensity and various expression activation patterns, illumination variations and small head pose variations are supported. Part of the work was completed before PAPUD. A theoretical and practical work about the performance<sup>2</sup> of various optical flow techniques (encoding the facial movement) for recognising expressions using convolutional deep neural network was conducted during PAPUD project and is now under review.

**Facial occlusions** We designed intelligent facial models in order to minimise the impact of occlusions. Initially, per expression and per occlusion type, weights were computed for the visible facial regions in order to minimise the occlusion impact. Later, fusion based techniques were explored in order to obtain per occlusion type models, regardless of the analysed expression. Reconstruction techniques of the occluded regions were also explored and published in 2019. We are now pursuing the occluded region reconstruction by means of deep convolutional auto-encoders. An arXiv version of the work is to be published shortly.

**Dissemination** Besides publishing in well known conferences (WACV – Core Rank A, CBMI – Core Rank B) and prestigious journals (TAC – IF 7.512, SPIC – IF 2.814, MTAP – IF 2.313), we also had the opportunity to present PAPUD in venues such as Research and Innovations Days organized by Lille University in 2018, 2019 and 2020. Around 200 master students each year visit our booth. Since, September 2020, we are involved in the Xperium exhibition organised between 2020 and 2022 at Liliad Learning Center where expression recognition prototypes linked to PAPUD activities are presented<sup>3</sup>.

#### 4.1.10 LORIA / University of Lorraine

In 2018, LORIA led WP1 “Scope and Requirements”. LORIA has participated to all three WP1 deliverables. LORIA has been particularly invested in deliverables D1.2 “Use Cases Definition” and D1.3 “State-of-the-Art”. For WP2 “Product Specifications”, LORIA has begun experimenting and implementing, through the use of PyTorch, deep learning models related to data provided by Bull-ATOS for Use Case 5 “Prescriptive maintenance for High Performance Computing”. For Use Case 4 “Behavior Analysis for Reverse Efficient Modeling”, LORIA has also begun experimenting and implementing, through the use of PyTorch, deep learning models related to user’s navigation paths when using a e-Service.

---

<sup>2</sup><https://arxiv.org/abs/1904.11592>

<sup>3</sup><https://lilliad.univ-lille.fr/actualites/xperium-kaleidoscope-limage-dans-toutes-sciences>

At the beginning of the project, we focused on supervised learning approaches for use cases 4 and 5. Supervision may come from two main sources: human manual annotations (HMA) and automatic annotations (AA). In use case 5 (HPC) supervision comes from automatic annotations, more precisely prediction of the next log line. Our objective is to detect specific events in the logs (e.g., fatal error messages); in use case 4 (BAREM), supervision was also automatic, driven by reconstructing the navigation traces of the user with a model derived from the family of auto-encoder models. The objective is to detect anomalies in the user behavior when navigating through an online application.

In **2019**, we continued our work on both use cases 4 and 5. In use case 5 (HPC), we mainly explored various types of deep learning models for generating the next line of log, in particular LSTM-based models, with and without attention, as well as Deep Averaging Networks. Note that although training equations are strictly supervised (generating the next log line), supervision is fully automatic and is available at no cost. Furthermore, we started developing the second stage of this HPC data processing pipeline, in the form of an unsupervised anomaly detection model, based on the state-of-the-art Deep-SVDD architecture.

In use case 4 (BAREM), we integrated our generative models on user traces within the video processing pipeline developed at University of Lille and the BAREM platform. This integration enables our model to leverage human manual annotations, in the form of user frustration levels. Therefore, we developed an additional supervised component to our model's architecture in order to optimize this new type of supervision.

In **2020**, we pursued our work on both use cases 4 and 5. In use case 5 (HPC), we mainly finalized our anomaly detection model based on the Deep-SVDD architecture and integrated it within the system logs processing pipeline developed at Bull-Atos. In particular, we designed a module to identify the key log sequence that characterizes a system failure, and that passed this log sequence to the Bull-Atos module that reconstructs the past tree of causal events that have led to this failure. We also started a new research activity on growing deep neural networks, that shall potentially better handle the continuous and "infinite" data stream coming from system logs. However, this research activity is still ongoing and is not mature enough to be integrated within the PAPUD platform yet. We will continue this research activity after the project.

In use case 4 (BAREM), we analyzed our supervised results on detecting user frustration, and identified a lack of meaningful correlation between the information contained in the user traces and the level of frustration. So, we switched to another NLP standard benchmark (the Reuters corpus) to validate our model and further proposed a novel fully unsupervised objective function to train our model. This new objective outperforms the previous state-of-the-art objective function published at ACL in 2019 by decreasing the number of errors made by the model by 50%, which is a large gain and sets a new state-of-the-art in the domain of text anomaly detection.

#### 4.1.11 Pertimm

Our intention was to find out how machine learning could be used for e-commerce website. We studied the whole value chain and look at what was the most interesting part apply the machine learning. This could concern :

- data cleaning and enriching
- putting product in categories (marketplace)
- improving merchandizing
- improving recommendation

In the first year of the project, we work on product categorization with the help of ContentSide : from a set of products and a three tiered categories set, and given textual data on each product, we calculated, using deep learning technics, the most probable categories for the product. The result was quite good with a best result of 98% a good categorization of products, on two datasets, one with 50 000 product and the second with about 1 million products.

In the second and third years of the project we concentrated on the recommendation systems for e-commerce websites. An e-commerce web site contains the following important data :

- A list of products and their characteristics
- A list of visitors and their characteristics
- A list of baskets of products bought by each visitor

Using those data, and working in collaboration with Setur, IMT and KocDigital, we designed and implemented a full recommendation system.

Using those data we demonstrated the use of artificial intelligence methods to create a recommendation system for clients of an e-commerce website. This use case was a joint collaboration with :

- Setur
- KocDigital
- IMT
- Pertimm

We demonstrated the use case using data from Setur which consisting of 31.000 products sold at Turkish airport duty free shops. We use also purchase data (baskets) to determine popularity of the products and calculate recommendation based upon a Deep Learning approach.

Recommendation are exposed to the user in two different ways:

- When showing a product, we propose several recommended products to buy based on this product
- When showing a result list using a search engine, we propose several recommended products, base on the 5 top products in the results list

As an outcome of the project, and since the demonstration proved interesting and successful, we decided to hire 2 AI person that are currently work on several recommendation systems that we will

integrate in our e-commerce platform:

- The first one is a vectorization representation on the sequential nature of user session on the e-commerce platforms. It is a baseline for session based recommender built on top of word2vec type embedding representation we use on transaction data and apply frequent item-set mining for developing recommendation on 'Frequently bought together' products based on the current product the user is viewing.

- We are also testing two recommendation models which are built on neural networks using Pytorch (deep) and Tensorflow, and one using Gensim library. Gensim word2vec model predicts at par with that of Tensorflow (0.04 second prediction time on average). PyTorch model predicts a bit slower but with more accurate results.

Those models will be included in our e-commerce platform, as well as our customers, website, to push forward more selling. We believe those will increase or competitiveness and revenues in the following years.

#### 4.1.12 ContentSide

ContentSide was responsible for work package : models definition and development.

Inside this work package, IMT was in charge of data preparation, ContentSide was in charge of algorithm designs and KoçSistem was in charge of integration.

ContentSide has created deep learning models for each use case :

UC1 E-commerce

- model for product categorization
- models for content recommendation and popularity detection

UC2 Call centers operations

- models for data categorization

UC3 human resources

- models for data classification

UC4 E-government

- models for sentiment/log analysis

UC5 Predictive maintenance

- model for prediction

Then, those models have been integrated into services by each partner responsible for the use-case. This collaborative project has been very beneficial for ContentSide as we integrated the model for product categorization into our SaaS solution ContentSide Semantic Platform.

A brochure presenting this solution can be downloaded from this address : <https://www.contentside.com/wp-content/uploads/2020/10/ContentSide-Semantic-Platform-Brochure-2020-EN.pdf>

Thanks to this integration, we can provide an important feature to online retailers : automatic product classification, very useful for online retailers having big catalogs (more than 1 million products) or for marketplaces that have to integrate catalogs coming from external sellers.

The selling of this solution started a few months ago, so today the revenue is not significant but our current selling pipeline is encouraging for the future.

#### 4.1.13 Setur

The role of Setur as a partner of the Profiling and Analysis Platform Using Deep Learning (PAPUD) project numbered 16037; Case provider and pilot process participant. Its main contribution is to contribute to usage scenarios, scenarios, requirements and dissemination activities and to provide data.

About providing data; Preliminary analyzes have been made on what data we can contribute. Since the amount of data is very important for the deep learning technology to give successful results, the evaluation was made by bringing all the data together. Deep learning studies and literature research have been done. Our R&D team was generally trained on artificial intelligence, machine learning, big data, and deep learning. The data were evaluated from this perspective. Parameters have been reviewed. After the analysis, 25 million rows of data were determined from the data set we worked on. Sample data set for Usage Scenario 1 has been shared with our project partners.

The deficiencies and inconsistencies in the data were evaluated. The data set to be shared after these analyzes was determined as follows;

- Products and categories
- Locations
- Customer id, age range, gender, marital status

Sample data set and parameters were shared with our partners for algorithm development studies. During the technical sessions of the face-to-face meetings, ideas were exchanged on usage scenarios. Contribution has been made to project documents in both national and international (ITEA) areas.

The following studies were carried out in order to make the data needs and related data available after the analysis with the partners, for the use of the Usage Scenario 1 stakeholders.

- Retrieving sales data of all stores in the past through Source Systems (25 million lines of receipt data)

- Identification of relevant store, product, customer dimensions and critical hierarchies for scenario studies for each dimension for detailing sales data.

- o Store Size: Zone-> Store Type (Check In, Check Out etc.) -> Store

- o Product Size: Category -> Brand -> Type1 -> Type2 -> Type3 -> Product

- o Customer Nationality Dimension: Geographic Region -> Country

- o Customer Size: Age, Gender

- Data integration studies regarding data that we can add externally besides sales data. Under this heading, a study has been carried out to obtain Passenger Data and Flight Information for Airports in the widest possible scope and to integrate them into the system. Information on the data received is as follows:

- o Daily Number of Passengers: The number of daily passengers passing through the relevant airport or border gate on the basis of shop type

- o Weekly Nationality Based Passenger Numbers: Weekly number of passengers passing through the relevant airport or border gate on the basis of shop type and nationality.

- o Flight Information: Receiving time, airline, airport and delay information of the flights taking place at the airport

- Integrating all data in a common structure and correcting the data that may be incomplete / empty / erroneous by performing data quality studies for each breakdown
- Sharing the integrated data with our project partners for Usage Scenario 1, informing about data breakdowns and contents
- Conducting joint studies to reveal applicable scenarios with data content to support our project partners
- We shared the data sets we received in our source system and with external integrations with our partners for algorithm development.

Contribution has been made to run and verify user scenarios on the platform to be realized by KoçSistem. It contributed by ensuring that the data provided for the tourism profiling / suggestion system usage scenario were prepared according to the platform and analysis needs.

For the e-commerce usage scenario, alternative scenario analyzes have been made with the data we have. Alternative scenarios were drawn and tested on what kind of results can be obtained through which data. Results after the test were interpreted. Due to our expertise in the tourism sector, our views have been formed on how the comments made as a result of deep learning will find a place in real commercial life. The data in the sales area was evaluated in terms of data analysis and presented to the



opinion of the sales representatives and the comments of our business partners on commercialization were obtained with the results. As improvements were made on the previously extracted data, the results were evaluated together with the partners, and support was provided to enrich the data set and increase the data quality in order for the algorithms to obtain more efficient results. For this, support was given with data analysis studies. In this image, age, country, etc. statistical values are shown.

Ay	Cins	Ülke		
<input checked="" type="checkbox"/> (All) <input checked="" type="checkbox"/> Ocak <input checked="" type="checkbox"/> Şubat <input checked="" type="checkbox"/> Mart <input checked="" type="checkbox"/> Nisan <input checked="" type="checkbox"/> Mayıs <input checked="" type="checkbox"/> Haziran <input checked="" type="checkbox"/> Temmuz <input checked="" type="checkbox"/> Ağustos <input checked="" type="checkbox"/> Eylül <input checked="" type="checkbox"/> Ekim <input checked="" type="checkbox"/> Kasım <input checked="" type="checkbox"/> Aralık	Search Cins	Search Ülke		
Ülke-Satılan Ürün		Ay-Ülke-Satılan Ürün		
Ülke	Prodcod	MusteriYas		
		P_S_Count		
		F M		
KANADA	KHL_GWP_U_BC	30	11	
KIBRIS RUM KESİMİ	OPI_MU_W_NM	37	9	
TURKIYE	KHL_GWP_U_BC	31	9	
KUVEYT	KHL_GWP_U_BC	23	8	
ÇİN	CLR_SC_W_FC	32	7	
HOLLANDA	KHL_GWP_U_BC	48	7	
TURKIYE	KHL_GWP_U_BC	30	7	
TURKIYE	KHL_GWP_U_BC	36	7	
TURKIYE	MAC_MU_W_FM	45	7	
BİRLEŞİK ARAB EMİRLİKLERİ	MAC_MU_W_LM	22	6	
BREZİLYA	MAC_MU_W_FM	49	6	
KATAR	TMG_GWP_W_GF	25	6	
KORE CUMHURİYETİ	KHL_GWP_U_BC	39	6	
TURKIYE	CHN_MU_W_EM	42	6	
TURKIYE	KHL_GWP_U_BC	50	6	
TURKIYE	KHL_SC_U_FC	43	6	
ALMANYA	MAC_MU_W_FM	29	5	
JAPONYA	KHL_GWP_U_BC	22	5	
KUVEYT	LOC_SC_W_BC	34	5	1
RUSYA	CLR_SC_W_FC	47	5	

In this example, products in a basket are entered on the left and suggested outputs for this basket are displayed on the right.

Ülke-Satılan Ürün	Ülke-MusteriCinsiyet-Satın Alınan	Ülke-MusteriCinsiyet-Yas-Satın Alınan
<b>Eklenecek Ürünler</b> Prodcod CAPT_CGR VILL_CGR		<b>Sepet Prod listesi</b> Prodcod P_SatisAnalizCount TSCN_CGR 17 CCREME_CGR 13 ALCP_CGR 10 BACKW_CGR 9 CHVS_WSKY 6 DANN_CGR 6 TOB_CHCLT 6 HNDLS_CGR 5 RTTSP_CHCLT 5 WRGLY_SKZ 4 HAVA_CGR 3 JW_WSKY 3 KING_CGR 3 KND_CHCLT 3 KNT_CIG 3 M&M_CHCLT 3 MHRS_CGR 3

Setur worked on the same data set with Pertimm and Koçsistem but on different usage scenarios. We worked with Pertimm with the demonstration of the usage scenario of the suggestion system specific to the search engine. With the use of the deep learning algorithm, Koçsistem contributed to the display of the scenario and the development of the scenarios created during the development process by interpreting them. In order to determine the application areas in commercial life in detail, the Sabiha Gökçen Airport duty free shops, which were selected as a pilot, were examined with the support of the business units and Koçsistem. Kiosks, digital displays, safes etc. The areas are determined as the areas where the suggestion system can be applied. The commercialization areas were determined. At the end of the project, as required by the implementation steps, the algorithms were tested with the business units Pertimm and Koçsistem. At the beginning of the project, the implementation was planned in Setur Duty Free shops at the airport, but unfortunately, the pilot implementation step could not be started, as the airports were closed due to the pandemic process.

#### 4.1.14 Softeam

Softeam, a French Software company, is the editor of the e-Citiz product which is a platform to design BPM processes. e-Citiz is used to design and deploy e-Services (aka Services proposed by web pages and workflow on internet). In eGovernment sector, citizens complain about painful processes but don't report bugs or enhancement to local authorities. PAPUD, with the BAREM Use case (Behavior Analysis for Reverse Efficient Modeling), aims to create a tool to analyze these problems and fix them. It will simplify and enhance the administrative procedures and then improve citizen satisfaction.

Softeam has actively worked with University of Lille and Lorraine (LORIA) to train deep learning models about frustration analysis and navigation analysis. One of the first tasks of these activities were log files anonymization because collected logs were real customers logs, with authorization from customer, but they needed to be anonymized. Then Softeam was involved in analysis approaches, how to get navigation, how to detect frustration, what may be the specific lines denoting frustration or navigation errors, ... This was the main research activities and it took a lot of time in the project.

On the other hand, Softeam led the WP3 about the deep learning platform and testbed and was one of the first partners to install the whole use case in the platform itself. Softeam, in addition to managing the WP3 deliverables, proposed modifications on the PAPUD platform to ease usage (like Docker).

Finally, Softeam has invested in building a component to display DL results and help e-Service designers to find bottlenecks and bugs.

From a dissemination point of view, the covid-19 crisis had a significant impact on dissemination activities since all of targeted meetings was cancelled or postponed after the end of the project.

From a business perspective, after the end of the project, we will try to go a step further on this BAREM module to reach a better stability and efficiency but the project has demonstrated the added value of the solution.

### 4.1.15 Turkgen

Today, Companies are spending a huge amount of money for increasing Customer Experience and Customer satisfaction all over the world. Data scientists are working on these hot topics that are analyzing the customers’ behaviors and conversations with call center employees to manage customer experience and satisfaction. Turkgen is working on analytics of data that is collected from call centers. There are 5 stages that are shown below to reach the valuable results.

- 1) Intent classification: This is the process of determining the reason for starting the dialogue.
- 2) Customer satisfaction measurement: It will be determined how the dialogue starts, progresses and ends with the help of sentiment analysis on different parts of the dialogue.
- 3) Agent performance measurement: Measuring criterion: Agent response times, Wrongly written word ratio, Change of mood values at the beginning and end of dialogues (customer satisfaction)
- 4) Entity Recognition: In order to get rid of unnecessary variations / details in the processing of dialogues, entities such as people, products, companies, brand names, prices, dates will be marked.
- 5) Correcting spelling mistakes: Word errors are too much. Correction needed to get rid of unnecessary variations.

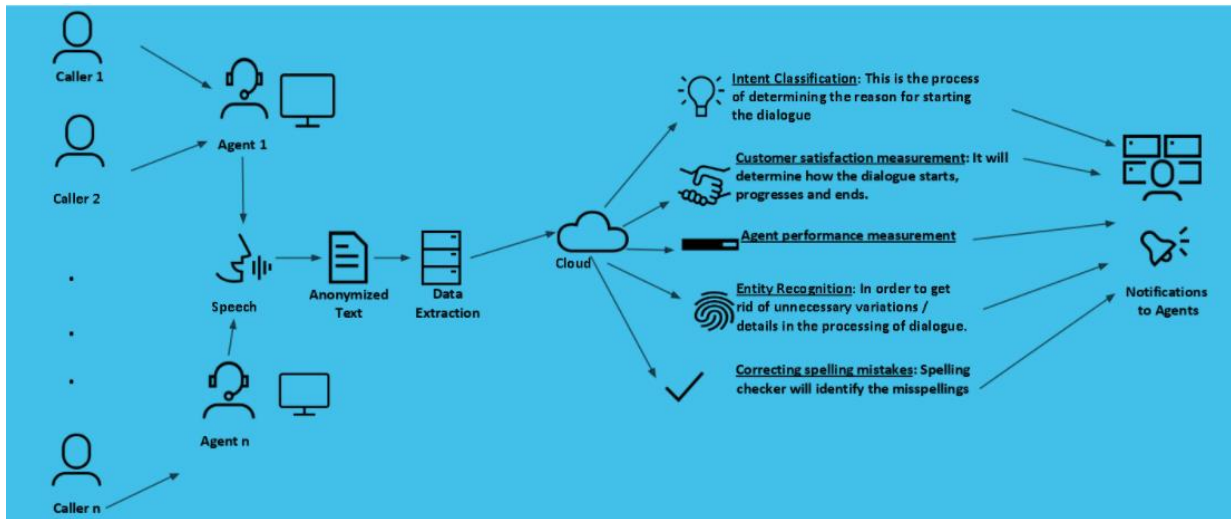


Table 1: High-Level Architecture

Each part that is mentioned above of the project is implemented in Python programming language using the algorithms that are coming from machine and deep learning text analytics literature. The Big Data environments that include MLLib Libraries, document-based NoSQL databases like MongoDB, Real Time Stream Analytic tools like Spark Streaming and infrastructure like Spark cluster used as a platform.

The use-case has been implemented as a proof of concept to one of the e-commerce company to measure the effectiveness of the activities. It has been showed in the study that the performance went up to %80 success by means of understanding the written text in the call centers.

Data which is related to Turkish use case that is about the call center has been gathered from an e-commerce company in discrete dialogues format. The data type includes customer agent dialogues in the call center as anonymized names of customers and agents, dialogue texts, message sending times.

Customer satisfaction is a term used to measures how products or services supplied by a company meet or surpass a customer's expectation. Measuring the customer satisfaction is one way help managers improve their business, but it is required an extreme analysis of different factors. Results of analytics will be evaluated according to customer behaviors that include Comparison of the metrics that are the frequency of calling the call center, the duration of the customer's dialogue etc. at the end of the project.

In the text preprocessing stage of the project; call center data moves to text documents. After taking text documents, the document is scanned to fix wrong writings using Levenshtein algorithm<sup>4</sup>. Documents are stored in MongoDB database and ElasticSearch. 80 percent of data is used for training the models that includes deep learning algorithms. After training the model, 20 percent of data is used to test model. Created model is used to classify the text data of each call center session. Sentiment analysis is also applied to customer sentences to predict customer satisfaction.

### Obtained Results

- Call center data has been extracted as document
- Wrong spelling mistakes have been fixed
- Models that has been created with DL Algorithms (Word2Vec, Doc2Vec, FastText), have been trained 80 percent of call center data and tested 20 percent of data.
- Sentimental analysis have been realized on the customer's conversation data.

#### 4.1.16 Turk Telekom

Türk Telekom (TT) is involved in developing the PAPUD framework using innovative big data and deep learning technologies for improving its own customer experience management systems. TT targets to exploit the flexible platform of PAPUD for improved service quality, churn analysis and customer satisfaction for retention and loyalty, TT targets to use project outcomes for ensuring increased cost efficiency and additional customer gains after the initialization of the platform. Türk Telekom predicts that the big data-based NLP methods in the project will be used in the contact center software market.

---

<sup>4</sup> Kogiso N, Watson LT, Gürdal Z, Haftka RT, Nagendra S. Design of composite laminates by a genetic algorithm with memory. Mech Compos Mater Struct 1994;1:95-117.

Project outputs will be used in customer call management systems where customer demands are received in the first stage and then to be used in customer loyalty and upsell / cross sell systems.

TT provided and used its data to train and use text based machine learning and deep learning models. Purpose of these algorithms is to automatize the extraction of topics from customer complaints raised against TT call center. The biggest two risk factors was the relatively small size of the data and the absence of labels in the data. Therefore Türk Telekom developed a data gathering tool and trained an unsupervised model to annotate the data. Besides from these two factors project architecture needed to be changed a few times to cover additional use cases.

TT performed R&D activities related to several NLP downstream tasks and developed deep learning and machine learning models. One example is topic extraction. For topic extraction, we are using unsupervised techniques including basic methods such as LDA and deep learning based methods. For supervised topic classification in addition to traditional machine learning based methods deep learning based methods also be used. We worked on real time multi-label classification of call center texts (e.g. assign complaint texts to existing topics and real time tracking and reporting of topics). We also conducted activities to gather and label three datasets used (two different data sets for the customer complaint and one data set for downloaded large amounts of data from Sikayetvar.com, major web portal in Turkey to track complaints).

In order to increase the performance of deep learning models we employ data science process which is an iterative one. The data and the AI model improved iteratively. In this process we also conduct significant activities in data preprocessing such as normalization, lemmatization, n-gram and collocation extraction. In order to obtain enough labeled data to start with deep learning modeling we use unsupervised topic modeling algorithms to assign multiple topic labels to the instances. After all data is tagged with this model we developed deep learning based multi-label text classifiers such as bi-LSTM. Along with multi-label topic classification we also developed several machine learning and deep learning models for sentiment classification and aspect-oriented sentiment analysis, all required significant effort to obtain data, labeling, preprocessing and modeling.

Another set of parallel activities are conducted for developing the POC of a commercial product. To make this product cloud ready we employed micro-service architecture and dockerization technology. To create micro services we have done developments with RESTFUL APIs and Flask. There are several functionalities can be accomplished by using this API and service interface such as adding new data to the system, new deep learning model training, listing all deep learning models with their performance metrics, classification of the entered text by the chosen model. NoSQL such as MongoDB and RDBMSs such as MySQL are also used in the implementation. Elasticsearch, Kibana, LogStash (ELK Stack) have been put into use especially for the advanced analytics dashboards for management and decision support purposes. As we employ iterative and evolutionary development approach, all the software along with the deep learning models are continuously improved during the project activities.

## 5 Partners efforts

We won't be able to provide the updated figures in terms of P/Y per partners since the project finishes officially the 15<sup>th</sup> of December and the partner's imputation won't be complete before this date.

We can use the present situation showing that **104.29 P/Y** were already spent for the project versus **103.12 P/Y** planned. There is already a slight overspending of 1,1% that will increase till the effective end of the project. We can observe in the ITEA site that most partners are inline with their planned efforts and the project itself also.

The main discrepancies that we have observed during the project life were due to difficulties to hire people, mainly for academic partners, at the beginning of the project. This situation has been caught up over the time. There were also discrepancies due to the reporting on the ITEA site, some partners were confused by the reporting in cumulative numbers in the semi-annual progress reports since all the other inputs were referring only to the current semester.

## 6 Impact and Exploitation

### 6.1 Impact on the employment

Here below are listed all the participants to the PAPUD project showing the impact on the employment for the project partners:

#### 6.1.1 4C

- Veerle Liebaut - project manager - 01/2018 - 09/2018
- Evelien Verbaenen - project manager - 09/2018 - 11/2019
- Sofie T'Kindt- project manager - 01/2018 - permanent
- Evelyn Caris - senior data scientist - 01/2018 - permanent
- Stijn Jaques- data scientist - 01/2018 - 12/2019
- Benoit Germonpré - data scientist - 01/2018 - 11/2019
- Quinten Claes - back-end developer - 01/2018 - permanent
- Pascal Zontrop - front-end developer - 01/2018 – permanent

#### 6.1.2 Beia

- George Suciu jr
- George Suciu sr
- Svetlana Segarceanu
- Stefan Stefanescu
- Ijaz Hussain
- Inge Cavat
- Mihaela Balanescu
- Christian Vasilescu

#### 6.1.3 Bull

All the Bull people listed below are permanent engineers working full time or partially on this project.

- Benoit Pelletier – DL Team leader and software architect
- Marc Platini- PhD “machine Learning applied to HPC systems failures analysis and prediction”
- Jean-François Roy – Cognitive data center leader
- Guillaume Porcher – HPC & Big Data R&D Engineer
- Cornel Crisan - PAPUD project Leader
- Laurence Londaits - PAPUD platform administrator
- Christine Martin - PAPUD platform administrator
- Nokeo Liankeo – PAPUD platform administrator
- Jeannine Pikus – HPC & Big Data R&D - FastML support Engineer

#### 6.1.4 Ericsson

- Sahra Sertalp – Project Leader , software developer - Permanent 2017-2020
- Tefvik Guler – Service delivery manager - Permanent – 2019-2020
- Alper Ozdogan – Software developer – permanent – 2018-2019

- Blilal Simsek - Software developer – permanent – 2018-2019
- Olgun Kaya - Software developer – permanent – 2018-2019
- Ruksen Inanir - Software developer – permanent – 2018-2019

#### 6.1.5 HI Iberia

- Antonio Fidel Pérez , Engineer – Permanent - 28/August/2017 to 22/June/2018
- Elena Muelas Cano – R&D engineer – Permanent - 01/January/2017 - 31/December/2020
- Fernando García SantaBarbara - Head of Development Department – Permanent - 01/January/2018 - 31/December/2020
- Inmaculada Luengo López - Head of R&D Department – Permanent - 01/January/2018 - 31/December/2020
- Fernando Cayón Valero - Developer - 01/Sept/2018 - 31/January/2019
- Juan Carlos Gracia Alique - Developer - 07/January/2018 - 31/December/2020
- Germán Fernández Sánchez – Developer - 01/January/2019 - 31/December/2019
- Luis Sánchez Acebedo - Senior Engineer - 01/01/2018 – 17/08/2018
- Raúl Santos de la Cámara – Senior engineer - 01/January/2018 - 31/December/2018
- Rubén Braojos Jerez - Engineer – Permanent - 01/January/2018 - 31/December/2018
- Tamara Martín Wanton - R&D Senior Engineer - 01/January/2018 - 31/December/2018
- Jose Javier Burguillo Sánchez - Developer - Permanent - 01/January/2019 - 31/December/2020
- Juan Miguel García Retamar - Developer – Permanent - 01/January/2019 - 31/December/2020
- Iñigo Goicolea Martínez – Developer – Permanent - 01/January/2019 - 31/December/2019
- José Manuel Grosso Venero – Developer - Permanent - 01/January/2019-31/December/2019
- Óscar Pérez – Developer – Permanent- 01/January/2020 - 31/December/2020
- Matilde Abascal Carmona - Developer - New Hiring – Permanent - 2/July/2018 - 31/January/2021

#### 6.1.6 Institut des Mines Telecom

- Marzieh Mozafari, PhD student, hired, from 03-05-2018 up to 03-05-2021
- Noel Crespi, professor at IMT and my supervisor, permanent, from the beginning of the project up to now.
- Samin Mohammadi, PhD graduate student, she was involved in the project from the beginning of the project up to 28-02-2019.

#### 6.1.7 Kocsistem

- Şebnem Güneş Söyler – Data Scientist/Technical Project Manager, permanent, 01.01.2018 – 31.12.2020
- Şeyda Akyıldırım - Data Scientist, permanent, 01.01.2018 – 31.12.2020
- Ahmet Çağatay Talay - Data Engineer, permanent, 01.04.2018 – 31.12.2018
- Ahmet Eray Bahtiyar - Data Engineer, permanent, 01.01.2018 – 31.12.2018
- Özgür Akarsu - Data Scientist, permanent, 01.07.2018 – 31.03.2019
- Fatih Yiğit – Developer, permanent, 01.01.2020 – 30.06.2020
- İsmail Burak Öksüzoğlu – Developer, permanent, 01.01.2019 – 31.12.2019
- Orhan Kınçak – Manager/Business Analyst, permanent, 01.01.2018 – 30.06.2018
- Hale Gezgen (supportive) - R&D Manager, permanent, 01.10.2019 – 31.12.2020
- Mehmet Önat (supportive) - R&D and IPR Manager, permanent, 01.01.2018 - 30.09.2019



#### 6.1.8 **KU Leuven**

- Geert Heyman - postdoc - 08/2018 -10/2018
- Bregt Verreet - postdoc - 10/2018 - 09/2020
- Sumam Francis - phd - 08/2020 - 08/2021

#### 6.1.9 **Lille University**

- Ioan-Marius Bilasco , Associate professor, Université de Lille
- José Mennesson, Associate professor, Université de Lille

Hired personel:

- Benjamin Allaert, research engineer, Sept 1st 2018 - April 30th 2019, left the project for a full time job as Lead Data Scientist at GenFit
- Amel Aissaoui, research engineer, July 15th 2019 - August 31th 2020, left the project for a full time position as Research Engineer at IMT
- Lille Douai Sofiane Mihoubi, research engineer, Sept 1st 2019 - August 31th 2020, left the project looking for a full time position as R&D
- Engineer Romain Belmonte, engineer and research engineer, Sept 1st 2019 - December 31th 2020

#### 6.1.10 **LORIA**

- Samuel Cruz-Lara, associate professor at the university of Lorraine: 25% since the beginning of the project
- Hubert Nourtel, expert engineer: Full time, since January, 2019
- Christophe Cerisara, head of the SYNALP research team of LORIA: 20% since the beginning of the project
- Paul Caillon, PhD student: 50% since Mai 23, 2019

#### 6.1.11 **Pertimm**

- Patrick Constant - President and founder of Pertimm -2017 -2020
- Sebastien Manfredini - engineer – 2017- 2019

#### 6.1.12 **ContentSide**

- Samir Amir – Engineer -PhD created the PAPUD project- 2017-2019
- Arnaud Dumont - President and founder of ContentSide -12/2019 -2020

#### 6.1.13 **Setur**

- Cenk Yusuf Ustabaş, R&D and Quality Process Leader, 19.03.2018-Working
- Müslim Erdal Şekerci, BI Manager, 11.06.2018-Working
- Osman Agah Arda, Product Process Leader, 16.10.2017-Working

#### 6.1.14 Softeam

- Laurent Goncalves, R&D project manager, permanent, 01/01/2018 – 31/12/2020
- Romain Thuillier, research engineer, permanent, 01/01/2018 – 31/12/2020
- Kais Chabouni, data scientist, permanent, 01/01/2018 – 31/12/2020
- Franck Verdiere, product manager, permanent, 01/01/2018 – 31/12/2020

#### 6.1.15 Turkgen

- Ibrahim Coskun - Solution architect – permanent
- Can Otur – Software developer – permanent
- Kadircan Turung - Software developer – permanent
- Betul Altiok - Software developer – permanent – 2019-2020
- Mete Aktas – Chief technology officer - permanent

#### 6.1.16 Turk Telekom

- Selami Ciftci
- Sezer Köylü Tokgöz
- Kubilay Aydın
- Engin Zeydan

## 6.2 Market and exploitable results

### 6.2.1 Target markets and main drivers

The targeted markets are defined in the Deliverable D1.1

- E-commerce – The B2B market is targeted by Pertimm with recommendation tools that will be enhance for their clients . Setur targets the airport shops and since the COVID19 oriented the tooling towards the internet market. ContentSide develops new technologies for the media analysis
- 4C uses deep learning models that enables companies to extract more insights from the interaction with the customer . Their actual customers are the car resellers. Ericsson with the Telecom data will will focus on building solutions for communications service providers based on Mobile network metrics and consumer service usage details.
- HI Iberia addresses the Human resources recruitment process with a Deep Learning based recommendation system
- With the BAREM uses Case the Softeam company addresses the e-government software market where Softeam proposes their application , products and e-services
- The HPC prescriptive maintenance Use case of Bull aims, with the output of the DL Training, to enhance the support applications, modify the support strategy and enhance the platforms conception to increase the platforms availability thus reducing the exploitation costs.

## 6.2.2 Exploitable results

The first exploitables results to be highlighted are the 3 Exploitables results selected within the top 8 achievements of the project :

- The recommendation system for e-commerce websites using deep learning systems, pushes products based on a list of products or only one product (Pertimm)
- The Detection of anomaly from system logs. It Calculate correlation between logs (e.g. a log in anomaly with the previous logs) (LORIA- Bull)
- The Recommendation system to support the Human Resources department in the candidates selection for a specific job offer. It provides also Human Resources department the salary variations along the years. (HI Iberia)

We can also mention :

- The improvement of the call centers operation (4C, KU Leuven, Ericsson, Beia, Turkgen, Turk Telekom)
- The product categorization with a good order (E-commerce Pertimm)
- The Deep learning popularity for E-commerce (IMT, Kocsistem, Pertimm)
- Improvement of e-government services (Softeam)

This list can be completed by the achievements related on the ITEA site. The one mentioned above are directly applicable to commercial markets.

## 6.3 Individual exploitation plans

This part doesn't include all the individual exploitation plans. Some partners developed more in depth these plans below, others mentioned the use of the project outcomes ins their partners work description above.

### 6.3.1 4C exploitation plan

- 4C will use the TellMi application as described above.
- Next, KU Leuven and 4C are developing a deep learning solution which automatically suggests possible responses to an incoming question of a customer, no matter which language he or she speaks. This will also be integrated in TellMi.

### 6.3.2 Bull exploitation plan

#### Exploitation objective

ATOS will actively work to leverage the technology research and development done during PAPUD project into Cognitive Datacenter product offering.

Indeed, Cognitive Datacenter is a software product used by IT administrators to predict future incidents, monitor the global sanity of complex systems and forecast key performance indicators. It already comes with a root cause analysis solution for metrics, but it is missing today a solution for offline log analysis. The problematic is especially relevant for HPC market because it is extremely difficult to navigate into millions of lines of logs when a many-nodes many-hours job is failing.

The go-to-market of this feature, repetitively asked by the customers, will be drastically accelerated thanks to the technology breakthrough brought by PAPUD project.

The results of the PAPUD platform development will enrich the Atos Codex with the project outcome and enhance the existing open and industrialized business analytics platform with HPDA features to tackle the most resource and performance demanding use cases.

### Positioning

The project outcome will be assessed from the Atos Codex prospective. The Atos Codex is a suite of business-driven analytics solutions and services. It supports public and private sector organizations in the quest to transform data into actionable business insight. It provides analytics consulting services and labs for rapid prototyping and proof-of-value. It offers both a repository of active use cases and affordable access to High Performance Computing resources. Atos Codex benefits from a worldwide network of 10,000 data scientists and analytics experts.

### What will be exploited

With his use case on HPC predictive maintenance Bull will address his HPC market (600 M€ turnover for Bull in 2019) and the much larger HPDA market.

### Impact

The maintenance costs are around 5% every year of the costs of the HPC clusters and the projected benefits of the project are, with the new tools based on Deep learning , to substantially reduce this maintenance costs and the penalties bound to clusters unavailabilities. The tools integrated in the Cognitive data center will help our support teams to save time and money.

#### 6.3.3 Ericsson exploitation plan

Ericsson exploitation plan of the project is to develop algorithms and data analytics methods to be used for Churn prevention in CRM and adding them to the CRM product.

#### 6.3.4 HI Iberia exploitation plan

- The business model envisaged for PAPUD is a **B2C business model**, that is, Recruitment Platform is sold to the customers (companies) to be used in the Human Resources Department.
- The idea for the business model is:
  - ❖ Initial deployment costs for the companies in order to adapt the PAPUD Human Resources Platform to the company requirements, i.e., CVs format, preferences of searching, etc.
  - ❖ Monthly fee based on different criteria as the usage of the service and the functionalities to be access.

### 6.3.5 **Pertimm exploitation plan**

As an outcome of the project, we are currently work on several recommendation systems that we will integrate in our e-commerce platform:

- The first one is a vectorization representation on the sequential nature of user session on the e-commerce platforms
- We are also testing two recommendation models which are built on neural networks (see details above in partner contribution)

Those models will be included in our e-commerce platform, as well as our customers, website, to push forward more selling. We believe those will increase or competitiveness and revenues in the following years.

### 6.3.6 **ContentSide**

This collaborative project has been very beneficial for ContentSide as we integrated the model for product categorization into our SaaS solution ContentSide Semantic Platform (see details above – in partner contribution)

The selling of this solution started a few months ago, so today the revenue is not significant but our current selling pipeline is encouraging for the future.

### 6.3.7 **Türk Telekom exploitation plan**

Türk Telekom has 15.7 million fixed access lines, 12.8 million broadband, 3.2 million TV and 23.1 million mobile subscribers as of September 30, 2020. With 321,000 km fiber connection, TT owns Turkey's largest fiber infrastructure in Turkey and owns more than 45,000 km of fiber infrastructure in about 20

countries outside of Turkey. Moreover, TT has the largest wireless backhaul infrastructure in Europe and operating more than 5,000 Wi-Fi hotspots access points in Turkey. In light of above facts, TT owns a big resource of data coming from these infrastructures and investments on big data is one of the key and most strategic item for TT. Hence, for exploitation of this data, servicing advanced text analytics using machine learning and deep learning for value added services in Customer Experience Management, personalized tariff/service/promotions for customer retention or cross-sell/up-sell opportunities, and advanced analytics dashboard for increased awareness of the sentiment and complaints will be provided.

## 6.4 Open source approach

We promoted the Open source with the preservation of the commercial goals of the private companies. Thus, only some of the developments will be provided in Open source. For example, Bull will introduce some developments in products (see exploitation plan) and some other tools will be put in Open source like the logflow tools.

Here below are the links to the githubs hosting some of the developments done during the PAPUD project, those githubs are provided by Bull, Lille University and the LORIA.

- <https://github.com/bds-ailab/logflow>
- [https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgitlab.u niv- lille.fr%2Fmarius.bilasco%2Fpapud\\_barem\\_frustration&data=02%7C01%7Ccornel.crisan%40atos.net%7C1a873c66b50d48eaa90408d854d62dc7%7C33440fc6b7c7412cbb730e70b0198d5a%7C0%7C1%7C637352626148412063&amp;sdata=vlx5IFl3u8TzFJ0g55KV6z2NkKxDOjyH%2FE%2FLxYgFeFo%3D&amp;reserved=0](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgitlab.u niv- lille.fr%2Fmarius.bilasco%2Fpapud_barem_frustration&data=02%7C01%7Ccornel.crisan%40atos.net%7C1a873c66b50d48eaa90408d854d62dc7%7C33440fc6b7c7412cbb730e70b0198d5a%7C0%7C1%7C637352626148412063&amp;sdata=vlx5IFl3u8TzFJ0g55KV6z2NkKxDOjyH%2FE%2FLxYgFeFo%3D&amp;reserved=0)
- [https://github.com/hnourtel/PAPUD\\_LogAnomalyDetection](https://github.com/hnourtel/PAPUD_LogAnomalyDetection)

The githubs access will be promoted by the academics in the conferences and workshops and the other partners in their dissemination activities (commercial events, workshops ..).

The access to the githubs will be monitored and the number of downloads evaluated to measure the adoption level of the proposed code and models.

## 6.5 Communication and dissemination activities

All these activities are detailed in the dissemination report (deliverable D5.2.2)

All those informations are also reported on the PAPUD Website created and maintained by the IMT.

<https://papud.wp.telecom-sudparis.eu/>

## 6.5.1 Communication and scientific dissemination

### 6.5.1.1 Scientific publications

#### 6.5.1.2 Publications accepted

27 publications (conferences and journal) have already been accepted, far above the objectives we had at the beginning of the project. This illustrates the important dissemination activity of the consortium members and the fact that the research done within the PAPUD project has been fruitful and innovative.

- Who Will Like the Post? A Case Study of Predicting Likers on Flickr - S. Mohammadi, R. Farahbakhsh, and N. Crespi - Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS). 2018, IEEE 2018
- Scrutinizing News Media Cooperation in Facebook and Twitter - P. Rajapaksha, R. Farabakhsh, N. Crespi – IEEE 2018
- Deep Dive on Politician Impersonating Accounts in Social Media - K. Zarei, R. Farahbakhsh, N. Crespi - IEEE ISCC 2019, Barcelona, Spain, June 2019
- Consensus Opinion Model in Online Social Networks based on Influential Users - A. Mohammadinejad, R. Farabakhsh, N. Crespi - IEEE Access, vol. 7, pp. 28436-28451
- User Reactions Prediction Using Embedding Features - S. Mohammadi, R. Farahbakhsh and N. Crespi - IEEE Global Communications Conference (Globecom), Abu Dhabi, UAE, December 2018
- OPIU: Opinion Propagation in Online Social Networks Using Influential Users Impact - A. Mohammadinejad, R. Farahbakhsh, N. Crespi - IEEE ICC, Kansas City, USA, May 2018
- CPU overheating characterization in HPC systems: a case study - Platini, Marc, et al. - IEEE/ACM 8th Workshop on Fault Tolerance for HPC at eXtreme Scale (FTXS). IEEE, 2018
- Impact of the face registration techniques on facial expressions recognition - Allaert Benjamin, Mennesson José, Bilasco Ioan Marius, Djeraba Chaabane - Signal Processing : Image Communication, EURASIP, Elsevier, 2018, 61, pp. 44-53
- Mastering Occlusions by Using Intelligent Facial Frameworks Based on the Propagation of Movement - Delphine Poux, Benjamin Allaert, José Mennesson, Nacim Ihaddadene, Ioan Marius Bilasco, Chaabane Djeraba - International Conference on Content-Based Multimedia Indexing (CBMI), Sep 2018, La Rochelle, France
- Etude de l'apport de la reconstruction des régions occultées du visage pour la reconnaissance des expressions - Delphine Poux, Benjamin Allaert, Nacim Ihaddadene, Ioan Marius Bilasco, Chaabane Djeraba - Coresa 2018, 12-14 novembre 2018, Poitiers, France
- Video-based Face Alignment with Local Motion Modeling - Belmonte Romain, Ihaddadene Nacim, Tirilly Pierre, Bilasco Ioan Marius and Djeraba Chaabane - Winter Conference on Applications of Computer Vision (WACV 2019), Hawaii
- Micro and Macro Facial Expression Recognition Using Advanced Local Motion Patterns - Allaert Benjamin, Bilasco Ioan Marius, Djeraba Chaabane - IEEE Transactions on Affective Computing, 10.1109/TAFFC.2019.2949559 – 2019
- Facial Expressions Analysis Under Occlusions Based on Specificities of Facial Motion Propagation - Poux, D., Allaert, B., Mennesson, J. et al. - Multimed Tools Appl (2020). <https://doi.org/10.1007/s11042-020-08993-5> - 2020

- Deep unsupervised system log monitoring - Hubert Nourtel, Christophe Cerisara and Samuel Cruz-Lara - PROFES2019 <https://profes2019.upc.edu/> 27-29 November 2019 Barcelona
- How much can Syntax help Sentence Compression? - Hoa Le, Christophe Cerisara, Claire Gardent - ICANN 2019, Sep 2019, Munich, Germany
- Multi-Lingual Dialogue Act Recognition with Deep Learning Methods - Jiří Martínek, Pavel Kral, Ladislav Lenc, Christophe Cerisara - Interspeech 2019, Sep 2019, Graz, Austria. (10.21437/Interspeech.2019-1691) 2019
- Quality of syntactic implication of RL-based sentence summarization - Hoa T. Le, Christophe Cerisara, Claire Gardent - Workshop on Engineering Dependable and Secure Machine Learning Systems of the AAAI-20 Conference on Artificial Intelligence (Oral presentation). New York, ABD – 2020
- Multi-task dialog act and sentiment recognition on Mastodon - Christophe Cerisara, Somayeh Jafaritazehjani, Adedayo Oluokun, Hoa Le - COLING, Aug 2018, Santa Fe, United States – 2018
- Do Convolutional Networks need to be Deep for Text Classification? - Hoa T. Le, Christophe Cerisara, Alexandre Denis - In AAAI Workshop on Affective Content Analysis – 2018
- Geolocation and Social Media for Enhanced Recruitment Campaigns - George Suciu, Cristina Bălăceanu, Adrian Pasat, Carmen Nădrag - 5th International Multidisciplinary Scientific Conference on SOCIAL SCIENCES & ARTS - SGEM 2018 <https://www.sgemsocial.org/>; Albena - Bulgaria; 24.08.2018 - 01.09.2018
- Facial Analysis Method for Pain Detection - Subea, O., & Suciu, G - In International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures (pp. 167-180). Springer, Cham March 2019
- A New Method to Help the Human Resources Staff to Find the Right Candidates, Based on Deep Learning - Vasilescu, C., Suciu, G., & Pasat, A - In The International Scientific Conference eLearning and Software for Education (Vol. 3, pp. 240-246). " Carol I" National Defence University – 2019
- EEG Signal Processing: Applying Deep Learning Methods to Identify and Classify Epilepsy Episodes - Suciu, G., & Dițu, M. C - In International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures (pp. 59-66). Springer, Cham March 2019
- Evaluation of Speech and Non-speech Classification Techniques for - Environmental Sounds Recognition - Elena OLTEANU, Inge GAVĂT, Svetlana SEGĂRCEANU - SISOM 2019 Conference on 16-17 May in Bucharest – 2019
- Speech Technology Today - Inge GAVĂT, Svetlana SEGĂRCEANU - SISOM 2019 Conference on 16-17 May in Bucharest – 2019
- Multifactor voice-based authentication system - Ceaparu, M., Toma, S. A., Segarceanu, S., Suciu, G., & Gavtat, I. - Conference in Telecommunications, Informatics, Energy and Management (TIEM) 2019, 12-14 September, Greece.

### 6.5.1.3 Publications under review

- CPU overheating prediction in HPC systems - Platini, Marc et al. - CCPE - <https://onlinelibrary.wiley.com/journal/15320634> - Upcoming



- LogFlow: Simplified Log Analysis for Large Scale Systems - Platini, Marc et al. - ICDCN - <http://www.icdcn2021.net/cfp.html> - Upcoming
- Peer to Peer WebRTC Multipoint Videoconference System - George Suci; Stefan Stefanescu; Cristian Beceanu; Marian Ceaparu - 10th Conference on Speech Technology and Human-Computer Dialogue - Upcoming
- A Non-Personalized Recommender System by Learning Product and BasketRepresentation - Savas Yıldırım, Şebnem Güneş Söyler, Özgür Akarsu - Turk J Elec Eng & Comp Sci - Upcoming
- What is the relationship between face alignment and facial expression recognition? - Romain Belmonte, Benjamin Allaert, Pierre Tirilly, Ioan Marius Bilasco, Chaabane Djeraba, Nicu Sebe - arXiv:1905.10784v1 – Upcoming
- Optical Flow Techniques for Facial Expression Analysis: Performance Evaluation and Improvements - Benjamin Allaert, Isaac Ronald Ward, Ioan Marius Bilasco, Chaabane Djeraba, Mohammed Bennamoun - arXiv:1904.11592 - Upcoming

#### 6.5.1.4 Workshops and events

Here below are listed the events attended by project partners:

1. EUREKA INNOVATION DAYS 2018, 22-24 May 2018 in Helsinki Finland – **BEIA**
2. IMWorld 2018, 3-4 October 2018 in ROMEXPO exhibition center Romania – **BEIA**
3. ICVL 2018 – **BEIA**
4. TMLSS 2018 – **BEIA**
5. COMM 2018 – **BEIA**
6. ECAI 2018 – **BEIA**
7. CeBIT 2018 – **BEIA**
8. SGEM 2018 – **BEIA**
9. SCEWC 2018 – **BEIA**
10. eLSE Conference 2019 – **BEIA**
11. SISOM 2019 – **BEIA**
12. CSCS22 Conference – **BEIA**
13. GloTS 2019 (17-21 June) – **BEIA**
14. Innovative Enterprise Week – **BEIA**
15. Marketing in Artificial Intelligence Era Conference 2019 – **BEIA**
16. SpeD Conference 2019

There are also upcoming events:

1. The international conference CBMI 2020 – **LILLE UNI**.  
CBMI postponed to June 2021. The conference will be held in Lille on June 28-30, 2021.
2. IPDPS20 - **BULL**



## 7 Conclusion

We had great ambitions at the beginning of the project hoping to build a Deep Learning platform that would fit for five very different domains. We realized a Deep Learning platform used by most of the partners for their models creation and training and then for the application part. It appeared during the project lifetime that there were many commonalities between the different application fields, each use case is set in a container allowing the portability, with Deep learning libraries, often common to all the uses cases. In a same way, each use case runs its application through web services.

The Deep learning models and applications proved their efficiency based on known algorithms and scientific novelties illustrated by the numerous publications and the validation process. This efficiency, demonstrated through the KPIs, gives to the different partners a competitive advantage for their respective markets. Some partners have already interesting commercial outcomes.

This project was also successful due to the close collaboration between the different project's partners that enriched through the exchanges their research and developments. This has been done under the umbrella of the ITEA promoting this collaboration.

The human capital with the numerous project's participants and the quality of those people is the base of this fruitful collaboration between partners that will foster future research and projects.