

## **Bachelor-Thesis**

zur Erlangung des akademischen Grades

Bachelor of Science (B. Sc.)

an der Hochschule für Technik und Wirtschaft des Saarlandes

im Studiengang Praktische Informatik

der Fakultät für Ingenieurwissenschaften

## **Einbindung der Mitarbeiter in Produktions- und Geschäftsprozesse in der digitalen Fabrik**

vorgelegt von

Mbonteh Ekuh Jude

betreut und begutachtet von

Prof. Dr Markus Esch

Saarbrücken , Tag. Januar 2018



# Selbständigkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ich erkläre hiermit weiterhin, dass die vorgelegte Arbeit zuvor weder von mir noch von einer anderen Person an dieser oder einer anderen Hochschule eingereicht wurde.

Darüber hinaus ist mir bekannt, dass die Unrichtigkeit dieser Erklärung eine Benotung der Arbeit mit der Note „nicht ausreichend“ zur Folge hat und einen Ausschluss von der Erbringung weiterer Prüfungsleistungen zur Folge haben kann.

*Saarbrücken , Tag. Januar 2018*

---

Mbonteh Ekuh Jude



# Abstract

This thesis objective is to develop a method to predict the salience of newspaper articles. We will be working on The New York Times Annotated Corpus and use the fact whether the article was on the Front Page or not as a proxy for salience. That is, we treat the problem as a binary classification problem with the classes Front-Page and non-Front-Page. Three key issues that need to be addressed as part of the thesis are.

- What is the suitable classification method to tackle the task ?
- What are effective features for it that can be easily extracted from the data-set.
- How can we deal with the imbalance training data?

Features will include standard words feature, but should also leverage the rich meta-data that is available (e.g., the topic of the article, its length, or the author). The thesis will include an experimental evaluation, studying the effectiveness of individual features or groups and features, and comparing the methods developed against reasonable baselines (e.g., a random classifier).



*A computer would deserve to be called intelligent,  
if it could deceive a human into believing,  
that it was a human [knuth:1976]*

—Alan Turing

## Danksagung

Hier können Sie Personen danken, die zum Erfolg der Arbeit beigetragen haben, beispielsweise Ihren Betreuern in der Firma, Ihren Professoren/Dozenten an der htw saar, Freunden, Familie usw.





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Problem Statement . . . . .	1
1.3	Solution Approach . . . . .	2
1.4	Outline . . . . .	2
<b>2</b>	<b>Foundation</b>	<b>5</b>
2.1	The Bosch Group . . . . .	5
2.1.1	title length . . . . .	5
2.1.2	article category . . . . .	5
2.1.3	word count . . . . .	5
2.2	SmARPro System . . . . .	5
2.3	Semantic Web Technologies . . . . .	5
2.3.1	RDF . . . . .	5
2.3.2	Classification Methods . . . . .	5
<b>3</b>	<b>Analysis</b>	<b>7</b>
3.1	Document Classification for Newspaper Articles . . . . .	7
<b>4</b>	<b>Concept and Design</b>	<b>9</b>
4.0.1	Weka . . . . .	9
4.0.2	Weka integration with Java (image) . . . . .	9
4.0.3	Data pre-processing in weka(sample article representation in arrf format) . . . . .	9
4.0.4	classification in weka(image) . . . . .	9
<b>5</b>	<b>Implementation</b>	<b>11</b>
<b>6</b>	<b>Evaluation</b>	<b>13</b>
6.1	SmARPro Platform . . . . .	13
6.1.1	Wearables . . . . .	13
6.1.2	INTEC Smart Device . . . . .	13
<b>7</b>	<b>Summary and Future Works</b>	<b>15</b>
7.1	Conclusion . . . . .	15
7.2	Future Work . . . . .	15
	<b>List of Figures</b>	<b>17</b>
	<b>List of Tables</b>	<b>17</b>
	<b>Listings</b>	<b>17</b>
<b>A</b>	<b>Appendix A</b>	<b>21</b>
<b>B</b>	<b>Appendix B</b>	<b>23</b>



# 1 Introduction

## 1.1 Background

In manufacturing, the internet of things and other innovative technology trends are re-defining the role of Information Technology (IT) in production. Multiple projects aiming at increasing data transparency as well as overall shop flow productivity are carried out year in year out. One of such project is Smart Assistance for Humans in Production (SmARPro). The SmARPro project is funded by the German Federal Minister of Education and Research (BMBF): following the increase of complexity in production operations due to innovative solutions for information driven manufacturing is a context-sensitive semantic data management platform for retrieving, storing, processing and presentation of data. The goal of SmARPro is to create a virtual twin of existing production and logistics system of a manufacturing plant with a special focus on human machine interaction via Augmented Reality. The Robert Bosch manufacturing plant in Homburg was used as a sample plant and the grinding machine "BAHMÜLLER UTRA twinner" as test machine. The SmARPro system consist of three main components: SmARPro Platform - a semantic data management platform for the context-sensitive storage, processing and provision of data and information. SmARPro platform creates a virtual image of the existing production and logistics system of the Bosch plant in Homburg, which is constantly updated. SmARPro SmartDevices - are Embedded Systems, which serve the connection of existing (in some cases also old) as well as new machines and plants to the SmARPro Platform. They are attached to the equipment in order to record sensor data (from dedicated sensors or existing control systems) and convert it into semantic models. Thus, any machines and systems can be enabled to Cyber-Physical Production Systems (CPPS). SmARPro Wearable - are electronic devices such as tablet PCs or Smart glasses that can be used as augmented reality-based assistance systems. They are used to assist employees in decision-making processes by obtaining, rendering and presenting information from the SmARPro Platform contextually (i.e., collaborator role, location, and task-related) information. They enable the exchange of information and the interaction between the employee and the machine

within a production system that can be integrated into the overall system by linking with the sensors developed in the research project. available, such as data glasses, tablet PCs and smartphones. They serve the exchange of information and the interaction between the employee and the machine. To support a future-oriented use of these main components, innovative methods and procedures for decentralized factory control are being developed. The term "Augmented Reality" describes the augmented reality that is achieved by adding additional information to the wearables and other aids. This is done, for example, by a smartphone, which records the reality with its camera, and additional information is displayed on the screen at prominent points.

## 1.2 Problem Statement

In a manufacturing factory such as the Bosch plant in Homburg, production and business processes generate huge amounts of data. These data can be used to increase the

## 1 Introduction

productivity of workers as well as machines. But due to lack of communication between workers and their production facilities, most of these data remain inaccessible and thus can't be put to used. Thanks to today's digitalization with industry 4.0, we can develop a system which creates a digital twin of a physical production plant. This digital twin is the SmARPro System. As mentioned earlier, this system comprises mainly of three key components: smart device, core platform and wearables. The SmARPro system comes with many challenges. The representation of context-based information on wearables for workers is one of them. This thesis deals with the real time presentation of machine data from the SmARPro platform on wearables for workers in a smart factory. To tackle this issue of presenting information on wearable, we first need to answer the following questions:

- How can a business process be reproduced and integrated on the SmARPro platform?
- How can human machine interaction be made possible within a business process chain?

The goal of this thesis is to answer the above questions. The sub section below provides a solution approach to tackle this challenge.

### 1.3 Solution Approach

With the rapid development of technology in the recent year plenty of innovation has been brought forward to aid facilitate our lives, one of this innovation is the SmARPro platform. using smart glasses which are connected to the SmARPro platform, machine operators can use augmented reality to analyze and solve complex problems quickly. the objective of the system is to assist machine operators with task such as machine repair in a scenario of a machine tool break down or mal function. Another task that come up after this machine repair by a machine operator is to provide a report for the machine break down after performing this repair

### 1.4 Outline

This subsection outlines the thesis structure. The chapters are structured as follows: Foundation - The foundation chapter will give a brief introduction of the industry partner where this thesis was conducted, then moves on to further explain the thesis keywords mentioned earlier in the introduction section namely: smart device, SmARPro platform and wearables. Finally, the chapter explains main concepts behind the SmARPro system, which include semantic web as well as Event-driven process (EPC). The keywords and main concepts will later serve as foundation work for the proceeding chapters. Analysis - The analysis chapter examines and explains sample relevant business use case scenario namely: Machine Tool Repair and Overall Equipment Effectiveness (OEE). It examines various steps of these business processes and pins points key aspect of human machine interaction within the respective process life cycles while seeking to provide the best interaction with the system. Concept and Design - This chapter builds on the analysis chapter. It provides sample prototypes of how the use cases mentioned earlier can be modelled and integrated in the SmARPro system while taking into consideration various aspect of Usability and human computer interaction. Implementation - In this chapter, we discuss implementations of concrete business processes in a typical manufacturing setting. This section explains how business processes can be modelled and reproduced on the SmARPro platform. Finally, we will demonstrate how this business process from

the platform can be presented on wearable devices. To help employees perform complex task such as machine tool repair. Evaluation - The evaluation chapter, discusses to what extend our solution approach was successful. We revisit the challenge of communication between machines, humans and their manufacturing plants. which we mentioned earlier in our problem statement and to what extend the work in this thesis could bridge this gap.

Summary and Future Works - The final chapter summaries and



## **2 Foundation**

In this chapter, fundamental concepts and terminologies required for understanding subsequent sections will be explained. We start with a brief background of the thesis industry partner: Robert Bosch GmbH, then move on to explain the main components of the SmARPro system, later on gives a brief explanation of semantic web concepts used in building the SmARPro platform. And finally, we discuss Event-driven Process Chain (EPC). We will later take a closer look at input from this chapter in more details in the analysis, design and implementation chapters. The content of this chapter is mainly based on the paper which describes the key components and technology of the SmARPro project.

### **2.1 The Bosch Group**

#### **2.1.1 title length**

#### **2.1.2 article category**

#### **2.1.3 word count**

### **2.2 SmARPro System**

### **2.3 Semantic Web Technologies**

#### **2.3.1 RDF**

#### **2.3.2 Classification Methods**

##### **2.3.2.1 Logistic Regression**

##### **2.3.2.2 Support Vector Machines**





## 3 Analysis

Plenty of related research has been conducted and still being made in the domain of Newspaper articles classification. In this section we will have a look at some valuable contributions, which are similar or closely related to our newspaper articles classification problem.

### 3.1 Document Classification for Newspaper Articles

In a research paper titled: *Document Classification for Newspaper Articles* based on worked conducted by [Ramdass2009] at Google, they aimed at investigating implementation techniques to automatically classifying articles from the MIT newspaper the Tech, into specific category (Opinion, Sport, News etc), by making use of natural language processing techniques for document classification. This techniques are predicted on the hypothesis that document or article in a specific sections have some distinguish natural language features. They found out salient features for classification ranged from word structure, word frequency, and natural language structure in each document. The data-set at their disposal consisted of a large archive of already classified documents, making it easy to apply supervised classification techniques. By randomly dividing their classified achieve into training and test data-set, they conduct experiments using various natural language feature set and statistical techniques, applying different classification algorithms such as Naive Bayes , Maximum Entropy, and comparing the performance in each case.



## 4 Concept and Design

Chapter four examines the tools, setups and various techniques used in tackling our article classification problem. we take a look at weka, our data mining tool, setting up weka, pre-processing our data in weka, and making use of weka built in machine learning algorithms to build a model classifier.

### 4.0.1 Weka

Waikato Environment for Knowledge Analysis (WEKA), is a java based open-source data mining tool, developed by the University of Waikato.(cit src). Weka is used in education, research, and industries, has also been implemented in Big Data technologies such as Hadoop, weka is a platform independent software, running on windows, Linux, and Mac see figure 1 for big picture of weka

### 4.0.2 Weka integration with Java (image)

### 4.0.3 Data pre-processing in weka(sample article representation in arrf format)

An Attribute Relation File Format (ARFF) , is the popularly used data format in weka, but weka can also read from Comma Separated Vaule(CSV) data format pros and cons of ARRF file and reading from databases. Taking a loot at sample arrf representaion of an article pre-processed data-set in figure x article-train-arff vs article-test.arff. We recall from chapter 2 that the classification process used training set to train the model to predict unseen data. so in our case , we are training , evaluating, and applying a classifier to classify newspaper articles into front and non front page articles( figure x )

### 4.0.4 classification in weka(image)



## 5 Implementation



# **6 Evaluation**

## **6.1 SmARPro Platform**

### **6.1.1 Wearables**

#### **6.1.1.1 Tablet PC**

#### **6.1.1.2 Smart Glasses**

### **6.1.2 INTEC Smart Device**





## **7 Summary and Future Works**

### **7.1 Conclusion**

**7.1.1**

**7.1.2**

### **7.2 Future Work**

**7.2.1**

**7.2.2**



**List of Figures**

**List of Tables**

**Listings**



# Appendix



## A Appendix A





## **B Appendix B**