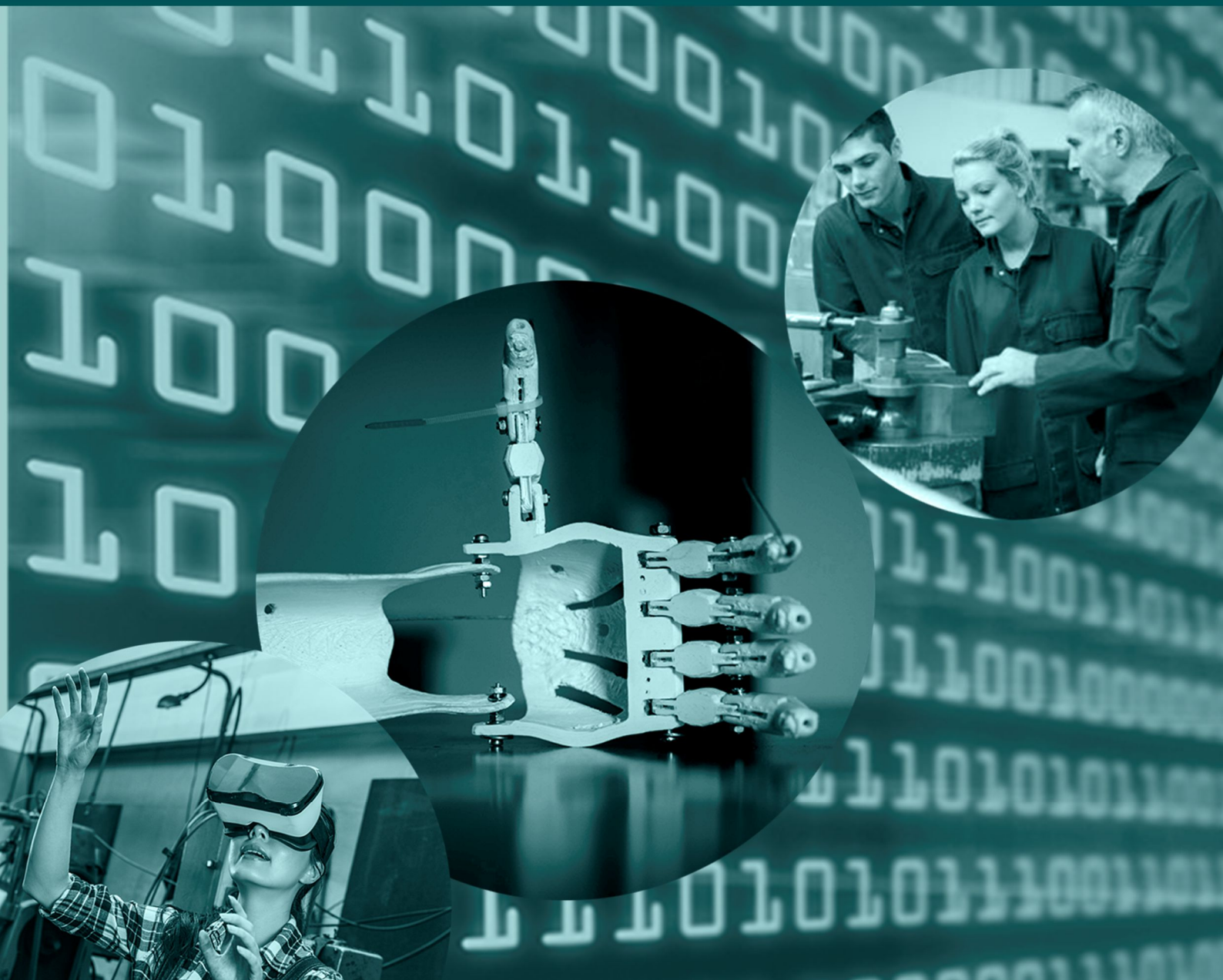
Big Social Data Analytics

REPORTING

INDUSTRIENS FOND FREMMER DANSK KONKURRENCEEVNE
The Danish Industry Foundation



Introduction

In 2015, at the time of the project start, the state of research and practice in the Danish industry with regard to big social data analytics was as stated below:

- Lack of software tools on the market that systematically combines sensitive business data with big social data for multiple business functions
- For companies exploring big social data analytics, they needed a tool chain of 5-8 separate tools that required separate technical and/or business skills
- Lack of Danish language support for social media analytics and many best-of-the-breed commercial tools did not support the Danish language
- Lack of large-scale commercial exploitations of research prototypes in big social data analytics
- Lack of business-oriented and company-focused big data courses or online courses and training resources for current and future employees of Danish companies

Based on the above, starting from 01-October-2015, this project addressed an urgent need for Danish companies to combine in-house business data with real-word big data by addressing the following research question:

How to extract meaningful facts, actionable insights, and valuable advantages from combining social media data (facebook, twitter, linked in, pinterest, instagram, blogs etc) with a company's in-house data (web analytics, eshop logs, CRM, ERP etc)?

The project aimed to provide three scientifically sound solutions to the research question posed above in order to meet the big data needs of Danish companies by overcoming the current limitations to the state-of-the-art:

- design, develop and evaluate business value of big data analytics tools in partner companies' test beds
- train employees at 1000 Danish companies to strategically use the big data analytics and measurable business value created
- train and certify about 45 Master students per year in the Data-in-Business Minor at CBS and measure the capacity-building of Danish companies in terms of new hires/roles in big data

Activities

The project consisted of four activities (Figure 1):

1. Big Data Collection & Processing
2. Predictive Models of Sales and Revenues
3. Dashboards for Social Media Monitoring
4. Training & Dissemination Workshops.

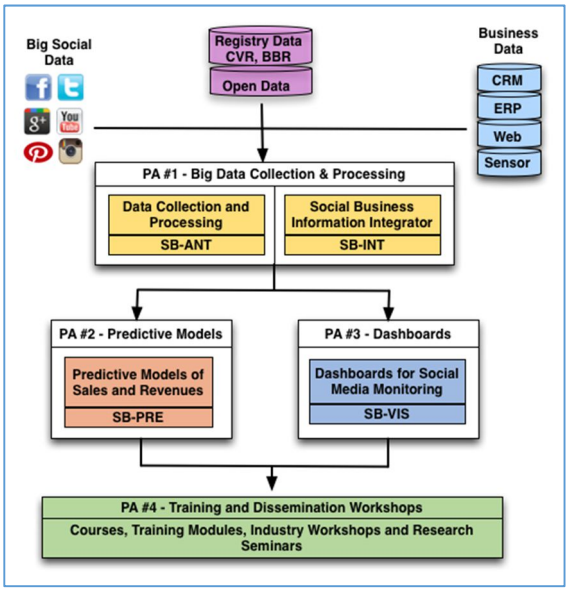


Figure 1: Project Activities

Deliverables

The project deliverables were:

- Software Suite
- Research Publications
- Big Data Courses
- Training Workshops

Software Suite

Three software tools (Figure 2) were designed, implemented, used and evaluated as part of the project's research and development activities.

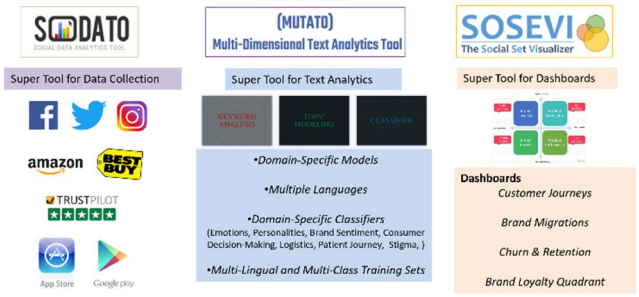


Figure 2: Software Suite

Social Data Analytics Tool (SODATO) supports the systematic collection, storage, retrieval and customized export of big social data sets from platforms such as Facebook, Twitter, Instagram, Amazon Reviews and Online Discussion Boards.

Multi-Dimensional Text Analytics Tool (MUTATO) supports unsupervised and supervised machine learning for multi-lingual and multi-class domain-specific text analytics applications such as brand keyword analysis, topic modelling and classifiers such as Consumer Decision-Making, Brand Sentiment, Brand Emotions etc.

Social Set Visualiser (SOSEVI) supports research-based set theoretical visualizations of big social data to explore users/consumers' brand associations, product interactions and social endorsements.

Additionally, the project designed, developed and evaluated a research prototype, *Social Information Integrator Tool (SOCIAL-IIT)*, that supports systematic injection of social information (behaviours and opinions) into digital products and services.

Research Publications

The software suite developed for and the big social data sets collected by the project activities supported the research activities of 2 Postdoctoral researchers, 6 PhD students, more than 60 Master thesis students and 3200s students in big data analytics courses at the Copenhagen Business School. In total this led to the publication of 2 books, 1 book chapter, 5 journal articles, and 44 conference papers with 2 best paper awards and 1 best paper nomination. Selected list is below:

- Østergaard Jacobsen, P., Bentzen, E., Bjerre, M., Bøggild, F., Hussain, A., Land, JF., Buus Lassen, N., Mukkamala, RR., Ringberg, T., & Vatrapu, R. (2015). Temperaturen på danske loyalitetsklubber anno 2015: Analyse & håndbog i arbejdet med loyalitet. Efficiens, Rungsted Kyst.
- Kunst, K., & Vatrapu, R. (2018). Understanding electronic word of behavior: conceptualization of the observable digital traces of consumers' behaviors. Electronic Markets.
- Siikanen, M., Baltakys, K., Kanniainen, J., Vatrapu, R., Mukkamala, R., & Hussain, A. (2018). Facebook Drives Behavior of Passive Households in Stock Markets. Finance Research Letters,
- Menon, K., Kärkkäinen, H., Jussila, J., Huhtamäki, J., Mukkamala, R. R., Lasrado, L., Vatrapu, R., & Hussain, A. (2018). Analysing the Role of Crowdfunding in Entrepreneurial Ecosystems: A Social Media Event Study of Two Competing Product Launches. Int. Journal of Entrepreneurship and Small Business.

Big Data Courses

Project activities supported research-based teaching and projects by more than 3200 students in the following big data analytics courses at the Copenhagen Business School.

- Social Media Management
- Big Data Analytics for Managers
- Minor in Data-in-Business
- Big Data Analytics
- Big Social Data Analytics
- Evidence-Based Management
- Visual Analytics
- Text Analytics
- Predictive Analytics

Tool Training Workshops

Tool training workshops were conducted in conjunction with the research-based teaching activities of the project. More than 50 tool training workshops in total were conducted for both best-of-the-breed commercial tools (Tableau, Power BI, IBM DSX, Alteryx, SAS etc.), open source tools (Python, R etc.) and research tools (SODATO, MUTATO, SOSEVI).

Topic	Tool
Data Mining-1	Altyrex
Data Mining-2	SQL
Visual Analytics-1	Tableau
Visual Analytics-2	Microsoft Power BI
Visual Analytics-3	SAS VA
Visual Analytics-4	IBM Cognos
Visual Analytics-5	SOSEVI
Text Analytics-1	Python
Text Analytics-2	MUTATO
Text Analytics-3	NLTK & Python
Predictive Analytics-1	Excel
Predictive Analytics-2	R
Predictive Analytics-3	SAS Studio
Predictive Analytics-4	SAS Enterprise Miner
Geospatial Analysis	Jupyter Notebooks
Big Data Analytics	IBM DSX
Facebook Analytics	SODATO
Image Analysis	scikit-learn

Effect

Over the course of the project big social data from more than 1500 organisations, brands, stores, and /or key influentials' have been collected and analysed as part of research, eductaion and training activities of the project. The main effects from the project were:

(a) proof of demonstration of the business value in big social data analytics for Danish organisations
(b) training of students and practitioners in big social data analytics for the capacity building of Danish organisations
(c) development of big social data analytics resources (training sets, domain-specific classifiers) for the Danish Language
(d) extensive contributions to the state-of-the-art in research on business data analytics with big social data
(e) capacity building in Danish companies in terms of a specialized staff function for big social data analytics
(f) single-loop big data analytics projects in Danish companies

Broader Industry Impact

The project contributed a research validated *Business Data Analytics Framework* to systematically combine in-house business data (sales, revenues, net promotor score, stock prices, marketing campaign effectiveness, web analytics) with external datasets on social media (Facebook, Twitter, Instagram, Online Reviews, and Discussion Forums).

Specifically, Danish companies can now explore and exploit big social data in their business development, innovation and growth process in the following ways:

- *Customer Segmentation*: Combining big social data with traditional marketing research data, Danish companies can create new customer segmentations based on advanced soft computing methods
- *Reputation Monitoring*: The project tools and results provide data-driven and evidence-based management methods for corporate reputation before, during and after social media crises
- *Brand Parameters*: Increasingly, the brand parameters such as brand identity, loyalty, coolness etc. are a function of what users and consumers are discussing on social media platforms and not just an outcome of traditional brand communications. The project provides innovative means of monitoring brand parameters through domain-specific text analytics

- *Predictive Analytics of KPIs*: The project provides case studies of using big social data for successful prediction of business performance outcomes such as sales, revenues, net promoter scores, and stock prices for a variety of products and services.
- *Digital Product and Service Innovation*: The project provides a design framework for injecting social information of two types, opinions and behaviours, which can be exploited in development of digital products and services.

Dissemination

More than 30 industry seminars and international conferences have been conducted. Public dissemination was conducted through the Danish Science Festival and public events at CBS, Danish Companies and international venues. Project activities at the Roskilde Festival's Big Data project was featured as an IBM Global Case and was widely covered in the Danish Media. The project supported the hosting and organization of the 9th International Conference on Social Media and Society at the Copenhagen Business School in July, 2019 and a number of ad hoc academic workshops and seminars on the theme of big social data analytics.

Going forward, after the end of the project in December 2018, the results of the project in terms of the software tools, training sets, research expertise, and industry training are embedded in and available from the Centre for Business Data Analytics (http://bda.cbs.dk/) and the new MSc in Data Science at the Copenhagen Business School. In order for Danish companies to ramp up their capacity-building of Business Data Science workers, educational activities of the project have been institutionalized into ad hoc courses, Cand-Merc.IT Minor in Data-in-Business and the new MSc in Data Science. Students in these courses and degree programs will continue to work with real-world datasets from Danish companies for their course projects, Bachelor projects and Master Theses and will serve as a dissemination vector for the advances in research at the Centre for Business Data Analytics at the Copenhagen Business School.

PROJEKTNAVN:
Big Data Analytics for Social Business

BEVILINGSMODTAGER:
Copenhagen Business School

PROJEKTANSVARLIG:
Kinga Szabo Christensen

MAIL:
Ravi Vatrapu

TELEFONNUMMER:
24 79 43 15

HOVEDNUMMER:
38 15 38 15

INDUSTRIENS FOND
FREMMER DANSK KONKURRENCEEVNE
The Danish Industry Foundation