THE BDSI NEWS

The Newsletter of the Behavioural Data Science Incubator

OVERVIEW

Introduction

Stephanie
van den Berg

AbhishtaAbhishta

Karin
Groothuis-Oudshoorn

Robert
Marinescu Muster

Martin Schmettow

Anna Machens

Karel Kroeze

Welcome to the very first BDSI newsletter.
BDSI has started last year with the aim to accelerate and inspire data driven research within BMS. We support researchers with statistical models, machine learning and simulation techniques. Since our start quite a lot has happened:

- The BDSI website has come online: bdsi.bms.utwente.nl
- We have started a series of webinars and workshops, to share some of our knowledge. It will be continued this year (see page 5). Videos of past webinars can be found on the webpage (here).
- BDSI has also created an online learning environment (OLE) to help with class projects, where students can run R code (see page 4).

- We have added two fulltime data scientists to the team and are currently supporting several research projects with data collection/web scraping, text analysis, machine learning advice and data visualization.
- This month we will also start a reading club. The first book we will read together with you is about statistical learning (see page 3).

Furthermore we have a section on tools and tips in this issue, showing a very simple way to use the magic of deep learning models for text analysis with python as well as a collection of interesting links comprising of a workshop for behavioral data science, a call for proposals for data science support from the eScience center and a hackathon to practice your data science skills.

THE BDSI BOOK CLUB

Reading club on Statistical Learning

We invite you to join the reading club on Applied Statistical Learning of the BDSI. Together we will read the book 'An introduction to Statistical Learning' from James et al. (link to a free copy).

This reading club is meant for all researchers from BMS interested in applying data driven methods to their research (in the future). This does definitely not imply that you are experienced yet.

We will meet every two weeks and each week we will prepare a chapter. The prerequisites for this club are basic knowledge of statistics. After we finished the book we want to apply the newly learned methods to your own data when possible.

You can join the reading club by sending an email to bdsi@utwente.nl with the topic: reading club.

The first meeting will be on Monday February 22th, 16:00 – 17:00.

If you have any questions you can e-mail to bdsi@utwente.nl

contact:
Karin Groothuis-Oudshoorn
Karel Kroeze

Statistical Learning

starts: 22/02/21
time: 16:00-17:00

THE BDSI OLE

BDSI Online Learning Environment

BDSI OLE

*Online Learning Environment
for learning R that Just Works*

BDSI has developed an online learning environment for supporting the transition to R for statistical analysis across BMS. The platform, powered by R Studio Server Pro, offers students and teachers a stable environment for writing and running R code, which is especially useful for students just making their first steps into R.

The environment is open to all BMS teachers. If you want to have more information, or start using the OLE for your course or practical tutorials in R, write us an email.

contact: Robert Marinescu-Muster

THE BDSI WORKSHOP

Using Internet based data collection for social science research

The Internet is a network of multiple networks spread across the world. In this digital age, empirical data collection on the network nodes or the world wide web can be useful in answering social science questions. Data mining over the Internet can also be used to study customer behavior, especially in a B2B setting.

In mid-March we are going to organize a workshop on the use of Internet measurements for social science research. Abhishta will be sharing how he uses internet measurements to estimate the economic impact of cyber attacks. The goal of this workshop is to discuss vantage points from which this data can be used for research at BMS.

Details of this event will follow. If you want to participate in the workshop, write an email to BDSI@utwente.nl with the subject "WEBINAR". For further questions contact: Abhishta Abhishta

next workshop
date: Thu 25/03/2021
talk: 12:15-13:00
hands-on: 13:00-13:45

Tools & Tips

NLP language models in three lines of code

In the area of natural language processing there is no way around transformer models. For a long time, these models have only been accessible for trained data scientists. However, Huggingface, a company that helps to facilitate working with transformers by maintaining python libraries, has recently released a new module, that makes the power of transformers accessible to almost everyone.

Pipelines

Now the power of many language models is available with just a few lines of code for tasks like:
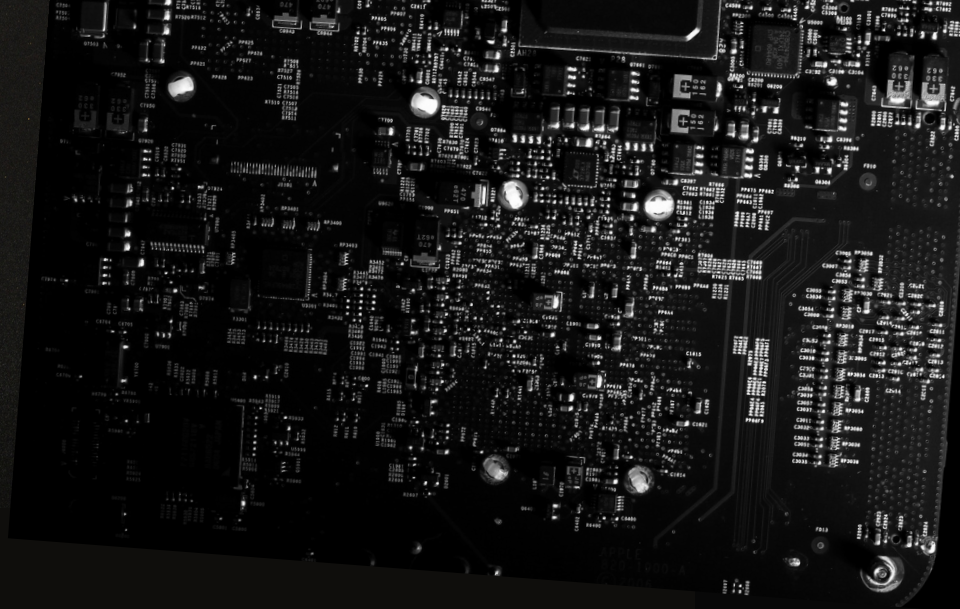
- Sentiment Analysis
- Named Entity Recognition,
- Summarization,
- Classification (Yahoo Topics)
- Translation
- Text generation

and much more...
what you'll need:
 python & a good internet connection for downloading the model.
Check it out here.

Code Example (python):

Sentiment Analysis

```
from transformers import pipeline
nlp = pipeline('sentiment-analysis')
nlp(" I really like BDSI")
```

>> [{'label': 'POSITIVE', 'score': 0.8309422135353088}]

Named Entity Recognition

```
nlp = pipeline('ner')
nlp("we are based in the Netherlands")
```

>>[{'word': 'Netherlands',
  'score': 0.9998524188995361,
  'entity': 'I-LOC',
  'index': 6,
  'start': 20,
  'end': 31}]

Or see various models here:
Zero-Shot Classification
Text Generation
all models

If you have questions you can email to
BDSI@utwente.nl
contact:
Anna Machens

From the web

Behavioral Data Science Workshop

The Statistical and Applied Mathematical Sciences Institute (SAMSI) held a virtual opening workshop for their new classes on data science and the behavioural sciences.
Talks include a presentation from Judea Pearl, who is known to a wider audience through his popular science book "The book of why", about *Deep understanding through Structural Equation Models (SEM)*.
Jukka-Pekka Onella talks about *Smartphone based digital phenotyping,*
David Lazer about the patterns of sharing fake news, Chris Bail about Political Polarization and many more. Videos of the talks can be seen online here.

Openssi 2021 - research support by the escience center

open call for small scale initiatives in data science

The Netherlands eScience Center offers training, consultancy and support from its Research Software Engineers for research projects for a maximum of 0.25 PYR. The call is organized around an intensive workshop (17-21 May 21) followed by several months of consultancy. **The deadline for submission is 1 March 2021, 12:00:00 (CET).** More information on the application process can be found on the website here.

Women in Datascience Datathon

The 4th Annual Women in Data Science (WiDS) Datathon is now open! This years datathon focuses on models to determine whether a patient admitted to an Intensive Care Unit (ICU) has been diagnosed with a particular type of diabetes. Teams need to consist of at least 50% women. The datathon is organized on kaggle. Team merging **deadline is Feburary 26, 2021**.
Get more information on WIDS website here.