European Network on Regional Labour Market Monitoring

The first seminar of the Big Data Knowledge Hub

Exploring new sources of Labour Market Intelligence: job seekers sentiment analysis

Claudiu Brandas and Ciprian Panzaru
Research Group on Social and Economic Complexity - SCOPE

Labor Market Analysis in the Big Data Era: advantages

- Large potential for analysis of labour market dynamics;
- Real-time analysis;
- Optimal labour market policies.

Labor Market Analysis in the Big Data Era: sources

Public online contents (job portals)	Social media (e.g., blogs, Twitter)	Web searches databases, search engines trends and analytics (google trends)	Operational metrics and other RTD (e.g., stock levels, school attendance, trainings, work permits, work contracts etc.)
Web content (news media, news articles, e-commerce, bibliographic databases)	Online information (web usage and content as a sensor of human intent, sentiments, perceptions (e.g., for career expectations)	Information actively produced or submitted by users through mobile phone-based surveys, apps	Information about purchases (in-store and online credit cards) and financial transfers

Labor Market Analysis in the Big Data Era: topics

- Labour demand and supply;
- Unemployment forecasting;
- Employment and earnings;
- Emergence of new occupations;
- New skills required on labour market;
- Skills needed by occupation;
- Post-graduate tracking;
- Career expectations.

Labor Market Analysis in the Big Data Era: examples and projects

LMI from OJA

- Information about location, occupation, educational level, salary, skills, sector, type of contract, working hours etc.
- The project Real-time labour market information on skill requirements: setting up the EU system for online vacancy analysis (CEDEFOP/CRISP).

LMI from IP location

- Estimate and predict short- and medium-labour migration flows through the Internet protocol (IP) addresses of website logins and sent e-mails.
- State et al. (2013) and Zagheni and Weber (2012) used IP addresses to map the geographic locations from where 43 million anonymized users sent e-mail messages within a given period.

LMI from social media

- Career histories provided by LinkedIn, labour mobility using geo-located posts on Twitter or using aggregated, anonymized data on Facebook users who list their professional status on their Facebook profile.
- Economicgraph.linkedin provide data about who is hiring, what jobs are available, and what skills are required in 180+ countries.

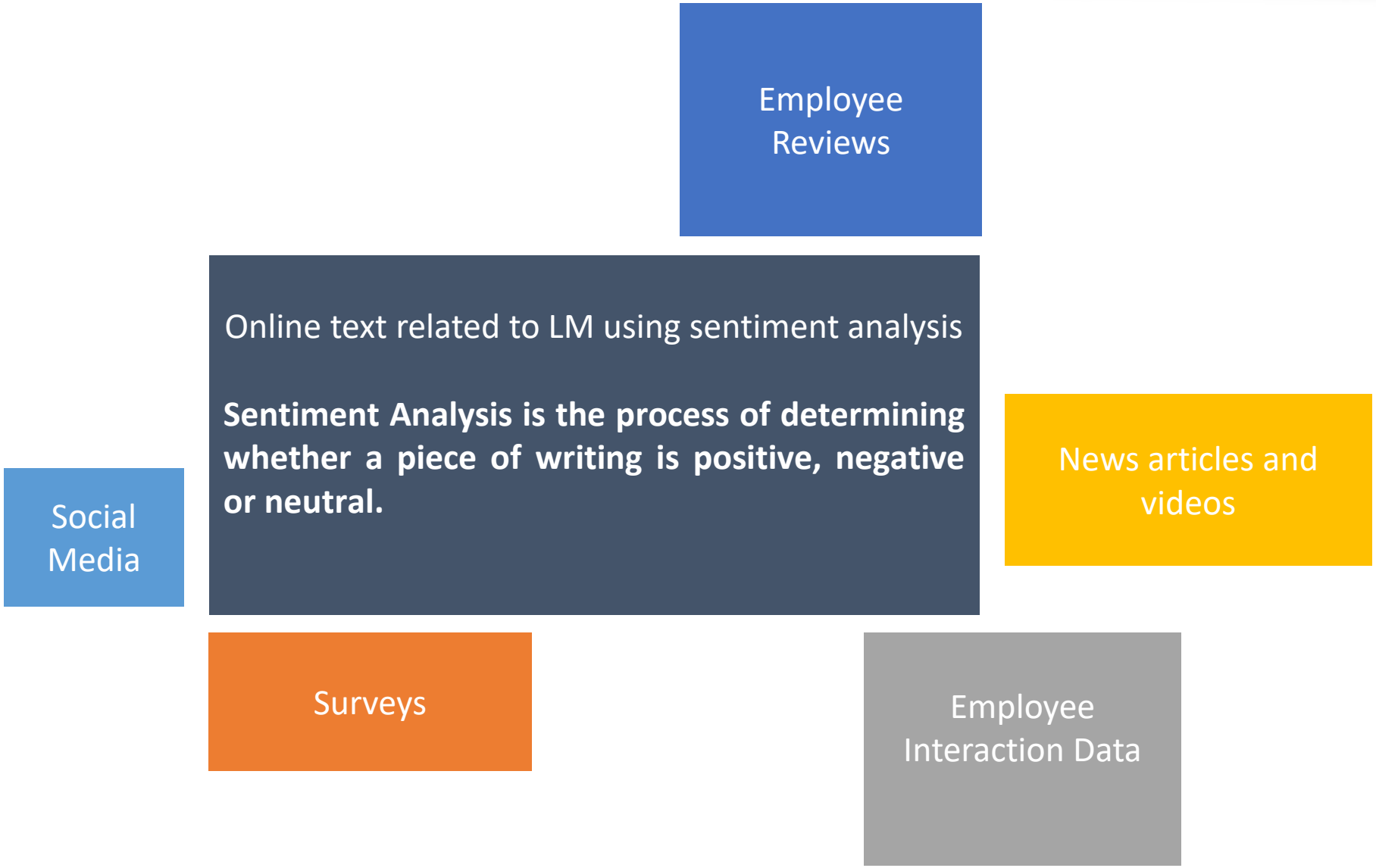
LMI from online search

- Estimations of labour mobility or unemployment using Google Trends (analyzing queries introduced to Google Search).
- Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018) combined official and Google Trends data to forecast the Italian youth unemployment rate.

LMI from mobile apps

- Mobile analytics involves measuring and analysing data generated by mobile applications. Data collected from the users, including profile information such as age and gender, location or search history can be analysed in Real-Time.
- Facebook Jobs, Glasdoor, Indeed

Exploring new sources of LMI: sentiment analysis



Sentiment analysis: definition

Sentiment analysis

- is a part of “Text Analytics”;
- identifies attitude (positive, negative, neutral) and emotion (joy, sadness, anger, pleasure, disgust, hate, surprise) from a message/comments (text, voice or video);
- using AI (Artificial Intelligence) algorithms for attitude and emotion classification. The most used is NLP (Natural Language Processing) algorithms such as: Support Vector Machines, Bayesian Networks, Deep Learning.

Sentiment analysis: common tools

Tools for SA:

- MS Azure Machine Learning – Text Analytics (*https://azure.microsoft.com/en-us/services/cognitive-services/text-analytics*);
- Google Natural Language API *(https://cloud.google.com/natural-language*)
- IBM Watson Natural Language Understanding Text Analysis; (*https://www.ibm.com/demos/live/natural-language-understanding/self-service/home*);
- Python with NLTK / TextBlob / Keras.

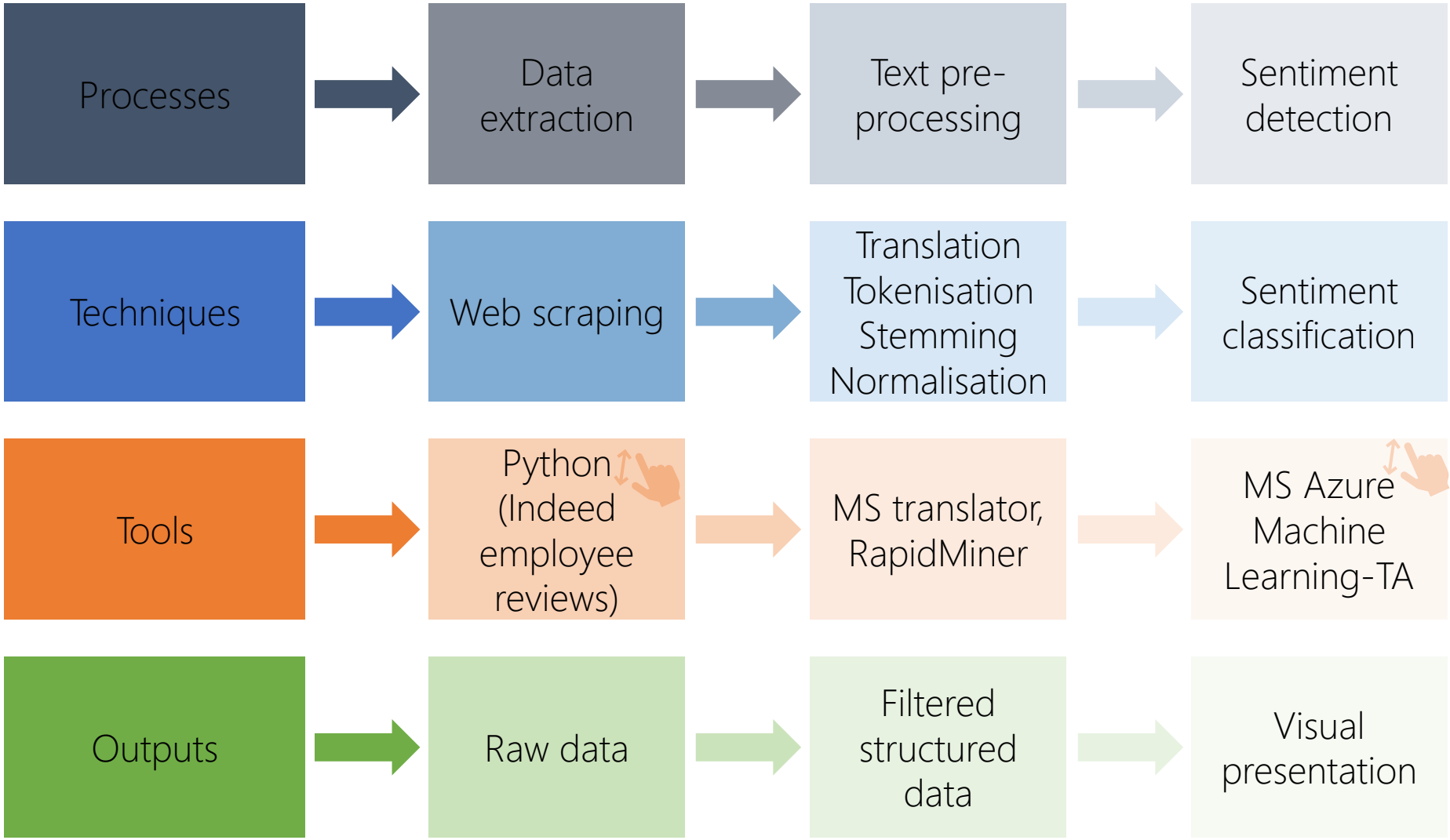
Sentiment analysis: MS Azure

Sentiment analysis using MS Azure uses Azure Cognitive Service for Language: a collection of machine learning and AI algorithms in the cloud for developing intelligent applications that involve written language.

Sentiment analysis: MS Azure

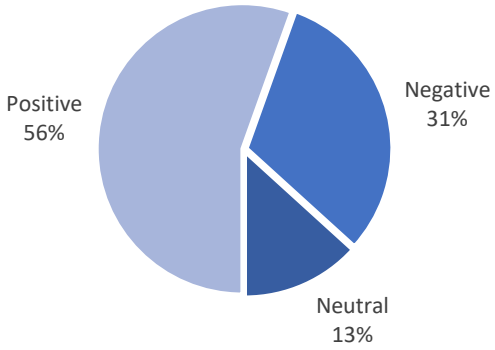
Sentiment Analysis using MS Azure Cognitive Service applies sentiment labels to text, which are returned at a sentence and document level, with a score for each. The labels are “positive”, “negative”, and “neutral” and the scores range from 1 to 0.

Job seekers sentiment analysis: research model

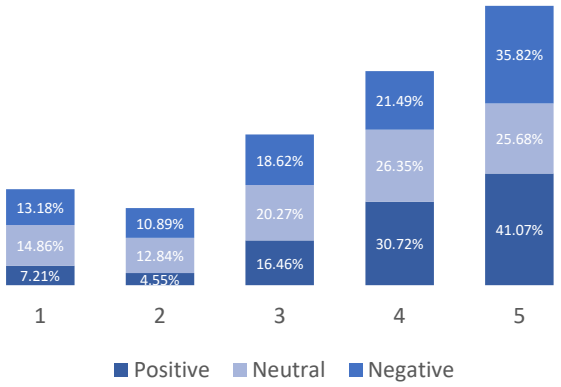


Job seekers sentiment analysis: results

Overall Sentiment

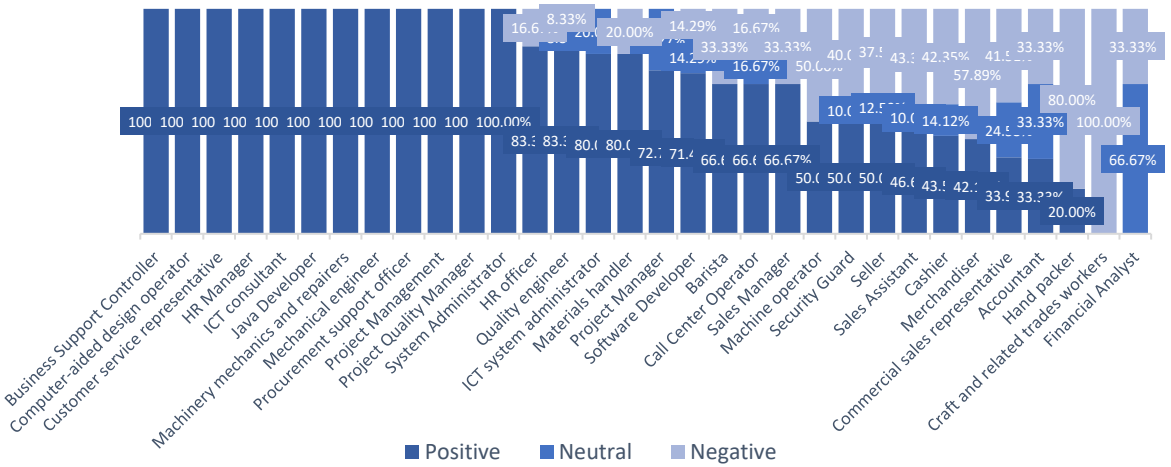


Sentiment by Ranking

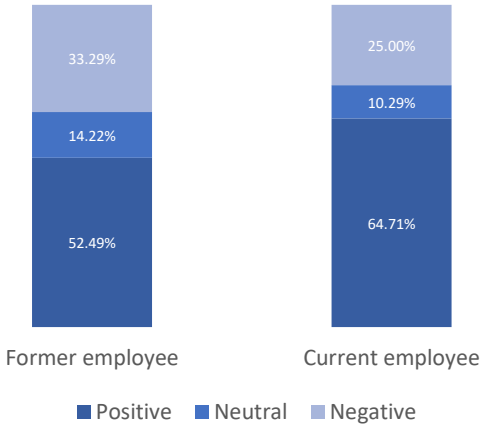


Keyword Cloud

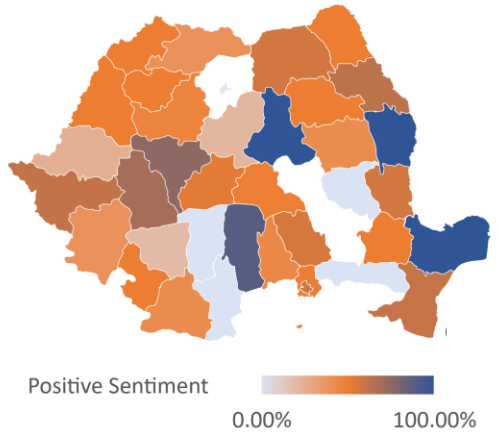
Sentiment by Occupation



Sentiment by Status

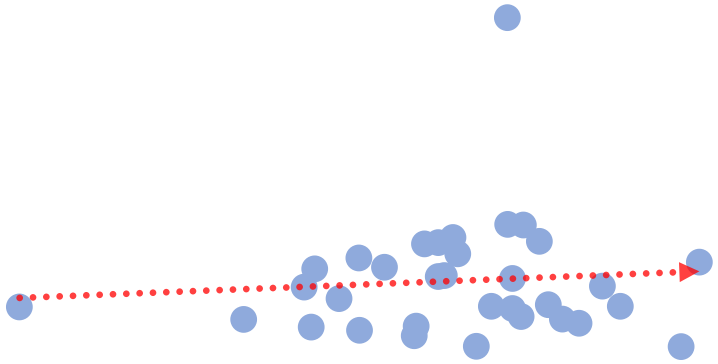


Sentiment by Regions



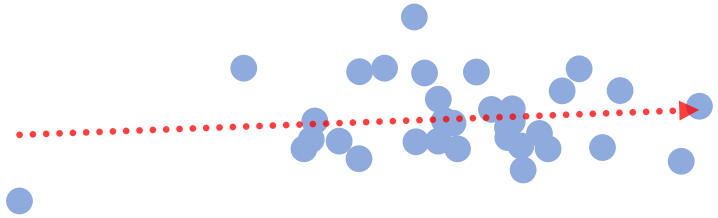
Job seekers sentiment analysis: correlation

Migration rate



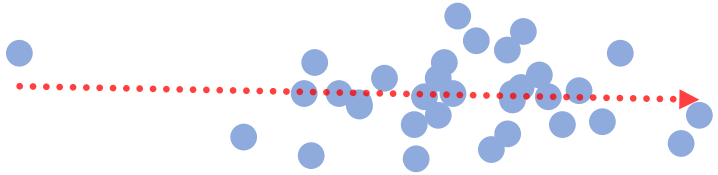
$R^2 = 0.0076$

Electoral participation



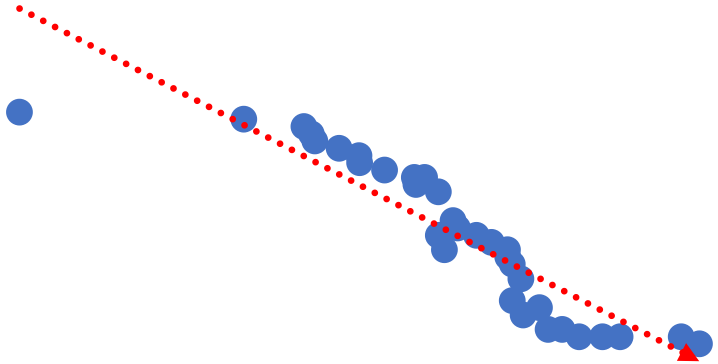
$R^2 = 0.0164$

Birth rate



$R^2 = 0.0052$

Unemployment rate



$R^2 = 0.8233$

Sentiment analysis: conclusions

For HR Analytics:

- Detect and understand employees feelings (e.g., concerning new workplace policies, changes in rewards and benefits, the workplace culture, etc. - improve hiring and recruiting practices, training and development needs);
- Contributes to data driven decision making process (e.g., identify triggers for positive and negative sentiment and identify where improvements could be made - help to build and maintain an engaged workforce and increase productivity).

Sentiment analysis: conclusions

For Policy Analytics:

- Contributes to data driven decision making process (develop various indicators that serve as “early warnings” on relevant topics - e.g., migration trends, unemployment etc.);
- Monitoring public policies and public investments (e.g., develop the public transport infrastructure for commuters).

Sentiment analysis: limitations

- Less procedures/methods to check data accuracy/data quality (genuine reviews? - what if these reviews are ALL scripted?, posted by trolls?);
- Sentiment analysis systems trained on English data exclusively;
- Automatic sentiment analysis of reviews using machine models are less accurate trained on review data related to labour market.

THANK YOU