Monitoring Adolescents' Distress using Social Web data as a Source: the InsideOut Project

Basili Roberto[†‡], **Bellomaria Valentina**[‡], **Bugge Niels J.**[⋆], **Croce Danilo**[†‡], **De Michele Francesco**[•], **Fiori Nastro Federico**[•], **Fiori Nastro Paolo**[•], **Michel Chantal**[⋆⋆], **Schmidt Stefanie J.**[⋆], **Schultze-Lutter Frauke**[⋆◦]

[†] University of Roma, Tor Vergata [‡] Reveal srl [⋆] University of Bern [⋆]University of Geneva [•] Sapienza University of Rome [◦] Heinrich-Heine University, Düsseldorf

{basili|croce}@info.uniroma2.it, bellomaria@revealsrl.it, niels.bugge@gmail.com

{chantal.michel|stefanie.schmidt|frauke.schultze-lutter}@kjp.unibe.ch

paolo.fiorinastro@uniroma1.it, {francescodemichele1981|federico.fiori.nastro }@gmail.com

Abstract

English. The role of Social Media in the psychological and social development of adolescents and young adults is increasingly important as it impacts on the quality of their interpersonal communication dynamics. The InsideOut project explores the possibility to use Social Web mining methodologies and technologies to collect information about adolescents' distress from their micro-blogging activities. The project is promoting a complex language processing workflow to approach the collection, enrichment and summarization of user generated contents over Twitter. This paper presents the general architecture of the InsideOut Web Platform and the resources produced by an integrated effort among computer science and mental health professionals.

Italiano. *Il ruolo dei Social Media nella crescita psicologica e sociale risulta essere sempre più importante poiché influisce sulla qualità e sulle dinamiche di comunicazioni interpersonali, specialmente riguardo le ultime generazioni. Il progetto InsideOut esplora la applicabilità di metodologie e tecnologie che consentono l'individuazione nel Web di evidenze riferibili a sorgenti di stress negli adolescenti. Il progetto propone un workflow di elaborazione linguistica in grado di gestire la raccolta, l'arricchimento e la sintesi dei contenuti generati dagli utenti su Twitter. Nel paper verrà presentata l'architettura generale della piattaforma Web InsideOut e le risorse che derivano dal lavoro congiunto di ricercatori provenienti dall'ambito informatico e medico.*

1 Introduction

Among adolescents, the use of Social Media, such as Twitter, Facebook or Instagram, has grown exponentially in the past years. This makes them a valuable source of information on the well-being of adolescents, but also concerning on their mental health. Mental disorders are the main cause of disability in adolescents and young adults (Gore et al., 2011), affecting an average of 10 to 20% of youth worldwide (Kieling et al., 2011). Thus, for the emerging complex relationship between the use of Social Media, mental health and well-being (Best et al., 2014), Social Media are a valuable source of information on the mental health and well-being of adolescents.

Social Media thus play an increasingly important role in the psychological and social development of adolescents as it impacts on the quality of their social interactions and networks. Any attempt to study and govern mental health in young communities (adolescents, students, interest groups) must take into account an effective and large scale methodology to monitor all the behaviors on the Web that exhibit and impact on mental habits, trends and social practices. The possibility of predicting writers demographics from their writings is an important research topic in the Computational Linguistic Community. In fact, the idea that a writer's style may reveal age, gender or other sociodemographic information has been also targeted in the "Plagiarism analysis, Authorship identification, and Near-duplicate detection" (PAN) (e.g., (Rangel et al., 2014; Rangel et al., 2015; Rangel et al., 2016)) or other experiences (Sulis et al., 2016) whose aim was to infer a user's gender, age, native language or personality traits, by analyzing the respective texts.

In this paper, the InsideOut project is presented. It explores the possibility to use Social Web min-

ing methodologies and technologies to collect information about adolescents' distress from their micro-blogging activities. The project is promoting a complex language processing workflow to approach the collection, enrichment of user generated contents on Twitter: messages written by a set of targeted community of users (e.g. from a school) are enriched with semantic metadata reflecting the expressed topics (e.g. social vs intimate relationships) and the attitude of the writers. The goal is to use this large scale evidence to support a comprehensive psychological characterization of adolescent communities and to pave the way towards effective applications of preventive and intervention efforts. The general architecture of the InsideOut Web Platform and the resources produced by an integrated effort of computer science specialists and mental health professionals will be presented. These data supported the exploratory evaluation where inter-annotation agreement scores and the performance over real data in the task of psychologically enriching user writings have been obtained.

In the rest of the paper, Section 2 describes the overall workflow underlying the InsideOut Platform. Section 3 describes the semantic models at the base of the semantic annotation process whose first result is the annotated corpus and the exploratory evaluation presented in Sections 4 and 5, respectively. Section 6 derives the conclusion.

2 The InsideOut Web Platfrom

The InsideOut Web Platform aims at supporting mental health studies concerning the causes of distress in adolescents. To this aim, a comprehensive service-oriented architecture has been designed and implemented to collect messages from Social Networks (such as Twitter) written by targeted communities of adolescents and enrich them with semantic information reflecting discussed topics and corresponding attitudes of the writers.

This enables specific kinds of queries and data aggregations, such as the pie chart shown in Figure 1, which summarizes the topics discussed by a community of users, e.g. concerning SCHOOL, FAMILY, or ALCHOOL AND DRUGS. By selecting a specific topic, such as SCHOOL, the system shows only those messages where the writer expresses a specific attitude, such as a DISTRESS. In the same Figure, the distressful messages concerning school are shown, such as "*Questa scuola fa schifo...*" ("*This school sucks...*") or "*Devo studiare.*" ("*I have to study.*").

In order to enable such queries the following services have been implemented:

Data collection services: services dealing with the extraction of data (messages/user information) from targeted social networks. These services are designed both to collect messages referring to a specific topic or hashtag, such as "*#maturità*" or messages exchanged between users belonging to specific *communities*, such as a members of a targeted school class. Among such services, we also implemented *Author Profiling* services that automatically determine the age of the writers (e.g. to filter adolescent's messages) but these specific services are out of the scope of this work.

Semantic annotation services: services dealing with the semantic annotation of gathered messages; once downloaded, they are automatically annotated with the semantic metadata described in the next section.

Storage services: services to store (possibly large-scale) collections of messages, communities and semantic metadata in NoSQL databases, implemented in MongoDB.

Reporting services GUI: services that aggregate messages, metadata and users to enable advanced report, such as shown in Figure 1.

3 Distress Characterization: The semantic modeling

In order to synthesize the amount of information made available on Social Media, we need to look at different semantic dimensions that can be associated with the writer's emotion, sentiment and mental status. Given that no direct diagnosis about mental health of an individual can be traced from or over one single message (but it is rather inspired by the observation of behaviors across temporal and social dimensions) we need to frame the mental state related information observable in Social Media within a comprehensive description of a subject.

So we decided to focus on the *experiential dimension* and start from the so-called **Life Event** dimension that expresses topics of interest and daily events in a young person's life. At the moment of writing, these have been discretized in eighteen different classes, as listed in Table 1. Each message can be assigned to one or more classes characterizing the possibly multiple topics

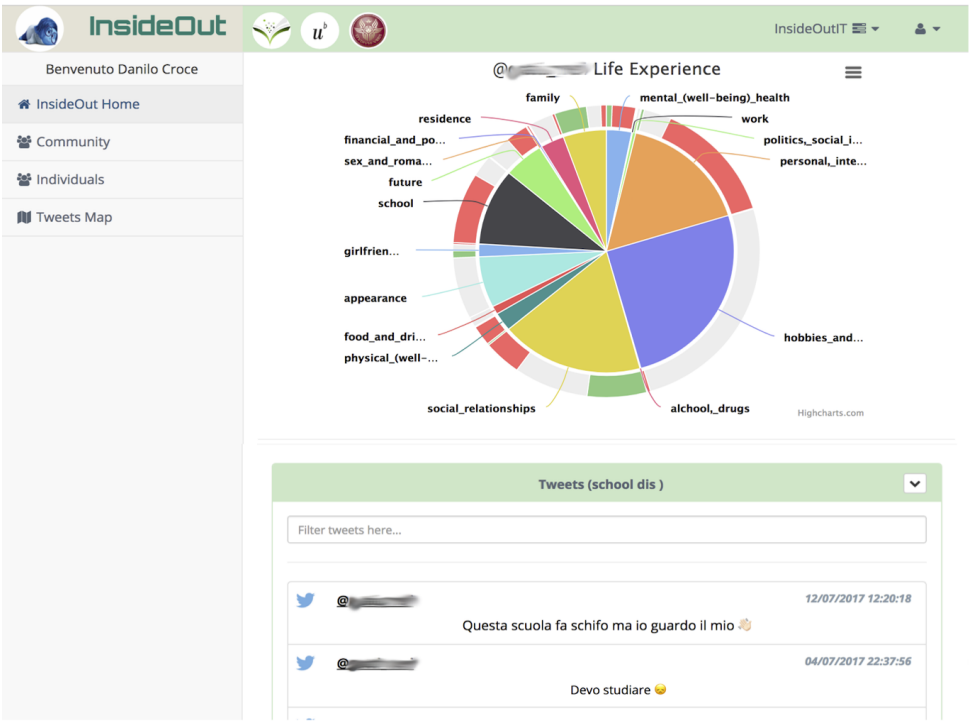

Figure 1: The InsideOut Interface

that can be mentioned in a message. For example, in the message "*Odio la scuola ma adoro i miei compagni*" ("*I hate school but I love my classmates*") the writer refers to the SCHOOL and SOCIAL RELATIONSHIP life events.

Moreover, a **Subjective** emotional dimension is targeted to capture the way the subject relates to the event in the micro-blog he writes, i.e., whether it is related to as a clearly positive or negative event, as a rather neutral statement, or in an ironic way. We referred to the traditional modeling for subjectivity analysis (Rosenthal et al., 2017; Barbieri et al., 2016), adopting POSITIVE, NEGATIVE and NEUTRAL classes; as an example "*Odio la scuola*" ("*I hate school*") is NEGATIVE, while "*Domani la scuola è chiusa*" ("*Tomorrow my school is closed.*") is NEUTRAL.

Finally, a further dimension called **Experience** tried to capture the writer's personal affect towards an event, e.g., whether it (*i*) is causing distress or other negative feelings such as anger or sadness, (*ii*) is regarded as helpful or causing positive feelings such as happiness or affection or (*iii*) is not associated with any perceivable emotional reaction (neutral). As an example, a school performance can be a positive experience if satisfactory for the teacher or the parents, thus being experienced as helpful by the writer, while it might be experienced as a negative event and as distressing when teacher's or parent's judgment is negative. It is worth noting that the Subjective and Experience dimension are nevertheless correlated, but they target different kinds of perception: the following message "*Mi sono rotto una gamba.*" ("*I broke my leg.*") can be considered DISTRESSFUL for the writer even if no agreement or rejection is made w.r.t. the event.

The information observable in a tweet is thus mapped into a set of three independent dimensions: (i) the type of Life Events le the message relates to (ii) the sentiment s of the event (POSITIVE, NEGATIVE, NEUTRAL) and (iii) experience-level e related to the event (among HELPFUL, DISTRESSFUL or NEUTRAL). For example, the tweet "*Quanto odio la mia classe... per fortuna mia sorella mi aiuta!*" ("*I hate my class so much... thankfully, my sister helps me!*") is assigned to the (le,s,e) triples: (SCHOOL, NEGATIVE, DISTRESSFUL) and (FAMILY, POSITIVE, HELPFUL).

4 The InsideOut Annotated Corpus

In the annotation process, annotators selected tweets written by adolescents (that have been previously manually validated) both in English and Italian and enriched them with triples (le, s, e), as discussed in the previous section. In the annota-

Table 1: Life Events description

Life Events	Definition
ALCHOOL, DRUGS	All actions and ideas involving the misuse of medications or the use of illegal drugs or alcohol.
APPEARANCE	All messages related to the physical appearance of the writer or of other people.
CRIME, ABUSE AND MOBBING	Thoughts, references and considerations directly connected to the world of crime or that express an attitude or an opinion of the adolescent towards that sphere.
FAMILY	Events involving or statements related to the family members, such as parental habits, relationships, generational clashes.
FINANCIAL AND POSSESSION	Event related to the financial status of the young person or his own family; needs of money for important needs or expectations; strongly perceived needs that strictly depend on the economic status and capability of the subject or his family.
FOOD AND DRINK	All actions and ideas involving food and drink (not alcohol).
FUTURE	Events or thoughts related to the perception the adolescent has about his own future.
GIRLFRIEND/BOYFRIEND PERSONAL RELATIONSHIP	(Usually strongly emotional) relationships based on sentimental and sexual attraction, involving gender aspects.
HOBBIES AND INTERESTS	All events, expectations or preferences evoked by entertainment related activities or personal interests (e.g. hobbies, fun, VIP) usually producing fun or connected with time-consuming helpful activities (e.g. games, TV, Social media, Celebrities).
MENTAL (WELL-BEING) HEALTH	Expressions related to mental well-being and to the health dimension but not related to physical aspects; this class includes sleep problems.
PERSONAL, INTERNAL STRESSORS, BELIEFS	General opinions, convictions or beliefs of the subject related to his own feelings and his personal sphere; general considerations regarding emotions, spirituality, stressors but not politics or social issues.
PHYSICAL (WELL-BEING) HEALTH	Thoughts, complains, considerations related to the physical health dimension, including conditions, nutrition, diseases, remedies and treatments.
POLITICS, SOCIAL ISSUES, ETC	All thoughts, considerations, reports regarding social, political and anthropological aspects of the close or general environment, as perceived by the young person.
RESIDENCE	Every perception about the locations where the subject lives or spends most of his time, including environmental aspects or weather.
SCHOOL	All events dominated by the school experience (comprehends social interactions IF only limited to school environment).
SEX AND ROMANCE	Events or experience specifically grounded at the sexual level, not including boyfriend-hood.
SOCIAL RELATIONSHIPS	All thoughts and events related to the relational dimension of the young people, but not involving the family, the criminal, the working/school and the boyfriend-hood dimension.
WORK	All events related to the relational dimension of the young people, caused or maintained alive by activities or dependencies based on the working condition of the subject of a member of his family.

Table 2: Corpus Statistics

Language	Italian	English
Number of tweets	2,037	1,072
- at least two annotators	1,074	1,072
- only one annotator	963	-
# annotators	4	4
# of (le, s, e) triples	2,517	2,811
Avg (le, s, e) for tweet	1.2	2.6
Avg token per tweet	16	15

Table 3: Inter-annotation agreement

Annotators Agreement - IT			
	Precision	**Recall**	**F1**
Life Event	85.76%	60.24%	70.76%
Sentiment	72.43%	50.88%	59.77%
Experience	74.28%	52.17%	61.29%
Annotators Agreement - EN			
	Precision	**Recall**	**F1**
Life Event	80.69%	56.17%	66.09%
Sentiment	63.77%	44.05%	51.99%
Experience	64.16%	44.35%	52.35%

tion process, each annotator starts by associating one or more le to a message[1] and, for each of them, the corresponding s and e must be provided. Each message was initially annotated by two annotators. After this first stage, the annotators in disagreement were asked to converge, in order to acquire a gold standard dataset. Only for Italian, we extended the dataset with a set of 963 messages that were annotated by only one annotator, without further refinements. The overall statistics of the dataset are shown in table 2.

In order to measure the complexity of the annotation process, we measured the inter-annotation agreement[2]. Given the possibility to associate more than one le to a message, we decided to measure the agreement in terms of Precision, Recall and F1, by considering the annotations confirmed after the agreement step as gold-standard and the initial annotations as measured annotations. Results are shown in Table 3. For the Sentiment and Experience dimension, we only focused on those messages sharing the same le. These agreement scores are quite low, confirming the difficulty of these kinds of analyses in Social Networks. The lowest score is Recall: it means that annotators generally assign a reasonable class, but it is very difficult to be exhaustive: as an example in the tweet "*Odio la gente che mastica rumorosamente. Mi innervosisce troppo!!!*" (*"I hate the people who chew loudly. It makes me very upset!!!"*) has been assigned to SOCIAL RELATIONSHIPS by an annotator while to PERSONAL, INTERNAL STRESSORS, BELIEFS by the other one. At the end, both were accepted and added to the gold standard. Agreement measured over the Italian messages is higher if compared with the English counterpart: one of the main reasons for this is due to

[1]Each annotator can associate zero, one or more les to a message.

[2]The inter-annotation agreement considered only messages annotated by at least two annotators.

Table 4: Results concerning the quantitative analysis of messages.

		Life Event			Sentiment	Experience
Lang.	**Tweets**	**Prec.**	**Rec.**	**F1**	**Accuracy**	**Accuracy**
En	1062	76.0%	31.3%	44.0%	62.8%	62.0%
It	1992	72.2%	47.2%	57.1%	67.8%	67.6%

the fact that Italian messages were annotated by native speakers, while English messages were annotated by German native speakers.

5 Exploratory Evaluation

In order to assess the applicability of the annotation process, we measured the quality of the system in the automatic recognition of Life Event (LE), Sentiment and Experience classes. We modeled this problem as a classification task and adopted the Support Vector Machine learning algorithm (Vapnik, 1995) in a One-VS-ALL schema, implemented within the Kernel-based Learning Platform (KeLP), presented in (Filice et al., 2015)[3]. We evaluated the three targeted dimensions of LE, Subjectivity and Experience separately[4] in a 10-Fold cross-validation schema: at each time a fold is selected as test set, while another set is the validation set used to estimated the SVM parameters. Each tweet is modeled by using the following feature representations: a *Bag-of-words* representation, *Bag-of-n-grams* (with $n = 2$ and $n = 3$) and a distributional representation based on Word Embedding (Mikolov et al., 2013) so that a message is the linear combination of its nouns, verbs, adjective and adverbs. For the LE classifier, we built a similar distributional representation of the eighteen LE definitions shown in Table 1: we introduced additional features in terms of the 18-dimensional vector containing the cosine similarity between the distributional representation of a tweet and the LE definitions. For Subjectivity and Experience, we added some specific features, modeling the presence of emoticons, punctuation marks (such as exclamation points), upper case words and elongated words. Moreover, we added features such as the length of the message (in terms of words and characters).

Regarding the LE dimension, we adopted a conservative strategy so that the system assigns a new LE to a message whereas the SVM classifier provides a positive confidence for the corresponding class while no LE is assigned, otherwise. Performance is thus measured in terms of Precision (the percentage of le correctly introduced by the system), Recall (the percentage of le from the oracle that have been correctly recovered) and F1 (the harmonic mean between Precision and Recall)[5]. Regarding the Subjective and Experience dimensions, once a le is known, the classifier is always requested to associate a message to the s and e labels, in order to generate consistent triples in the form (le, s, e). Being a multi-classification schema were the classifier always outputs a class, Precision is always equal to Recall[6], as well to the F1. In order to avoid redundancy, only one measure is reported and it is referred as Accuracy as it also corresponds to the percentage of messages correctly associated to the gold-standard label.

Preliminary results are shown in Table 4, both for English and Italian. Regarding the LE dimension, the adopted strategy results in a Precision higher than 70%, but at in a lower Recall. We believe this is mainly due to the reduced size of the dataset: it is even more relevant for English where only a 31% of Recall was detected. This number is consistently higher for the Italian dataset, where almost the double of examples is in fact provided and almost half of the tweets were only annotated by one person, thus reducing the odds for differences in annotations. Anyway, these results are consistently higher with respect to a baseline: the correct LE classification given the random selection from 18 classes would achieve a F1 no higher than 3%; if we require two correct classifications, in line with the average le per tweet shown in Table 2, this baseline drops to 0.3%. Moreover, it is worth noting that the adopted conservative strategy has been adopted to have a higher precision: since we are able to collect a huge amount of messages from social network, we can afford to lose

[3] Available at `www.kelp-ml.org`.

[4] When considering Subjectivity and Experience, a gold standard Life event is assumed.

[5] Since a message could be associated to multiple le the evaluation is not message-based but annotation-based.

[6] It may be the case that the LE classifier produces a number of les different from the number of the ones provided in the gold-standard. As a consequence, when evaluating this specific classifier, each message potentially introduces a different number of false positives and false negatives, so Precision and Recall will diverge.

some messages (often characterize by too little information in very short messages) instead of introducing too many noisy meta-data in the overall workflow. Results concerning sentiment are generally consistent with respect to international benchmark in English (Rosenthal et al., 2017) or in Italian (Barbieri et al., 2016) where almost all systems achieved an Accuracy between 60% and 65% (even using larger datasets). Overall, this result seems to be significant, as in line with the first outcome of the inter-annotation agreement. However, a further analysis is required to adopt more complex models for classification of such short messages, such as more complex kernels (Agarwal et al., 2011) or deep methods (Kim, 2014).

6 Conclusions

This paper summarizes the InsideOut project where the possibility to use Social Web mining methodologies and technologies to gather evidence about the adolescents' mental distress.The semantic model defined here and the annotated resource pave the way to a long-term joint research between computer science specialists and mental health professionals. The outcomes suggest the applicability of the devised methodology to larger communities and different languages. Since the system is currently active over Twitter, the final version of the paper will discuss about 5 months of continuous monitoring outcomes towards Italian and English speaking communities, with interesting evidences about the future of our project as a novel and ambitious Social Computational Science application.

References

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, LSM '11, pages 30–38, Stroudsburg, PA, USA.

Francesco Barbieri, Valerio Basile, Danilo Croce, Malvina Nissim, Nicole Novielli, and Viviana Patti. 2016. Overview of the evalita 2016 sentiment polarity classification task. In *Proceedings Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016), Napoli, Italy, December 5-7, 2016.*

Paul Best, Roger Manktelow, and Brian Taylor. 2014. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review*, 41:27 – 36.

Simone Filice, Giuseppe Castellucci, Danilo Croce, and Roberto Basili. 2015. Kelp: a kernel-based learning platform for natural language processing. In *Proceedings of ACL: System Demonstrations*, Beijing, China, July.

Fiona M Gore, Paul JN Bloem, George C Patton, Jane Ferguson, Vronique Joseph, Carolyn Coffey, Susan M Sawyer, and Colin D Mathers. 2011. Global burden of disease in young people aged 1024 years: a systematic analysis. *The Lancet*, 377(9783):2093 – 2102.

Christian Kieling, Helen Baker-Henningham, Myron Belfer, Gabriella Conti, Ilgi Ertem, Olayinka Omigbodun, Luis Augusto Rohde, Shoba Srinath, Nurper Ulkuer, and Atif Rahman. 2011. Child and adolescent mental health worldwide: evidence for action. *The Lancet*, 378(9801):1515–1525.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar*, pages 1746–1751.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Francisco Rangel, Paolo Rosso, Irina Chugur, Martin Potthast, Martin Trenkmann, Benno Stein, Ben Verhoeven, and Walter Daelemans. 2014. Overview of the 2nd author profiling task at pan 2014. In *CLEF evaluation labs and workshop*, pages 898–927.

Francisco Rangel, Fabio Celli, Paolo Rosso, Martin Potthast, Benno Stein, and Walter Daelemans. 2015. Overview of the 3rd author profiling task at pan 2015. In *CLEF 2015 Evaluation Labs and Workshop Working Notes Papers*, pages 1–8.

Francisco Rangel, Paolo Rosso, Ben Verhoeven, Walter Daelemans, Martin Potthast, and Benno Stein. 2016. Overview of the 4th author profiling task at pan 2016: cross-genre evaluations. *Working Notes Papers of the CLEF*.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation*, SemEval '17, Vancouver, Canada, August. Association for Computational Linguistics.

Emilio Sulis, Cristina Bosco, Viviana Patti, Mirko Lai, Delia Irazú Hernández Farías, Letizia Mencarini, Michele Mozzachiodi, and Daniele Vignoli. 2016. Subjective well-being and social media. A semantically annotated twitter corpus on fertility and parenthood. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016), Napoli, Italy, December 5-7, 2016.*

V Vapnik. 1995. The nature of statistical learning theory.