

White Paper
Italtel Data Science Solution

Value-producing innovations around data

June 2018

Contents

The Use Case and Proof of Concept 4
Technical Solution 7
Immediate Applicability 11
Some Results 12
PoC evolution and other possible use cases 15
Components Proven Evidences 16
Final thoughts 16

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

ABSTRACT

The initial idea was to propose innovation on data-driven business critical processes of an Operator and demonstrate the value that a data science approach can bring in a telco-specific context.

The idea gradually took shape and consistency and now a path from a simple prototype for fast evidence to a solution in place has been marked out working with real data, blending and tuning multiple software components (already widely proven in other contexts) and in the end, the comparison with SP's real problems. We meant the delivery of something tangible, this technical document provides readers with an at-a-glance overview of the challenges and results.

"EARLY INTERCEPTION OF CUSTOMERS' DISSATISFACTION"

A Data Science Use Case for Vodafone

Author

Silvio Gerli: Prof. of "Complex Data Analysis" at Statistics and quantitative methods department, University of Milan (Bicocca) and C.E.O. at Smartstat, a data science leading company, Italtel Partner for Data Science Solutions.

Contributors

Several people participated in this experimentation. We would like to thank Vodafone team for their contribution in both PoC definition and realization.

We like to point out that it was a team of experts (Data scientists, statistic experts, developers, big data experts, data analysts) with complementary skill-set in a context in which the word data scientist need to associate with the strong application domain expertise. It is here that the strong web analytics skills have joined the specific telco background and has resulted in a shared and appreciable result.

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

The Use Case and Proof of Concept

CSPs have access to a wide range of data, in more abundance than ever, from their own systems, at different speed and in different formats and this can be powerful. But to efficiently deliver a superior customer experience, they need to enrich their traditional telco data silos using external unstructured data such as CSPs' webpages, blogs, forums and social networks.
In this area, Italtel has developed a specific use case named "Early Interception of Customers' Dissatisfaction". The proof of concept proposed to Vodafone is based on this use case as well.

The scope of the use case is to early detect the *customer dissatisfaction "wave"* by crawling web data sources before Customer Care is heavily impacted. So the technical team could:

1. Intervene sooner on a problem
2. Have immediately more details to face with the problems.

The main idea is to continuously gather and analyze data that can be publicly got from the web in order to know what the subscribers said in real time: coverage problems or other service issues, similarly billing or contact issues, together with any sort of administrative issues, just to give some examples.

The following web data sources can be accessed by means of the crawler:

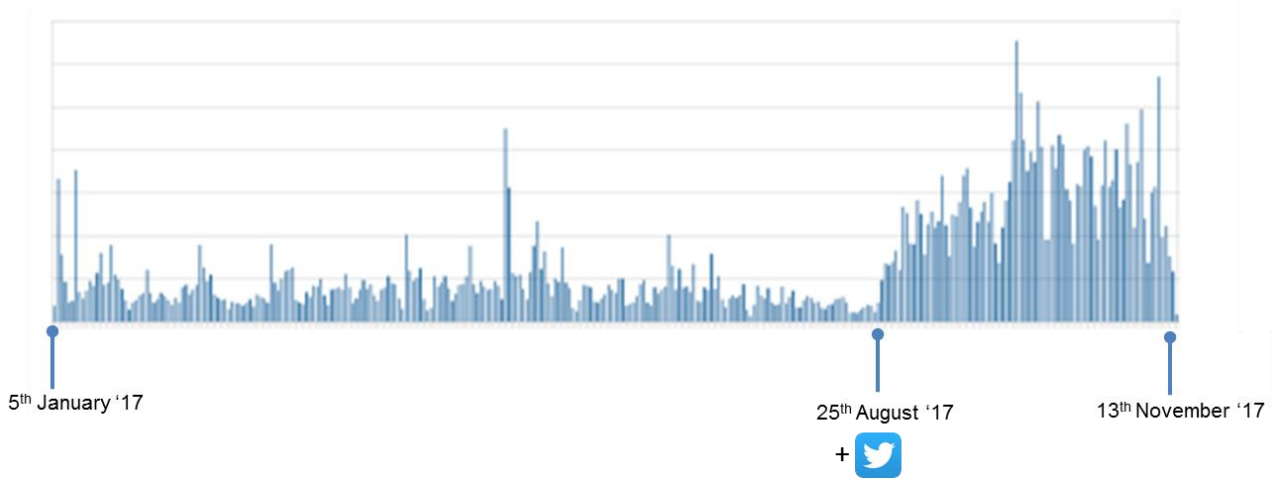
1. Vodafone's web site
2. Telecom Forums
3. Telecom Blogs
4. Social Networks
5. Editorial web sites
6. Web pages

and the following other data sources by means of a webapp:

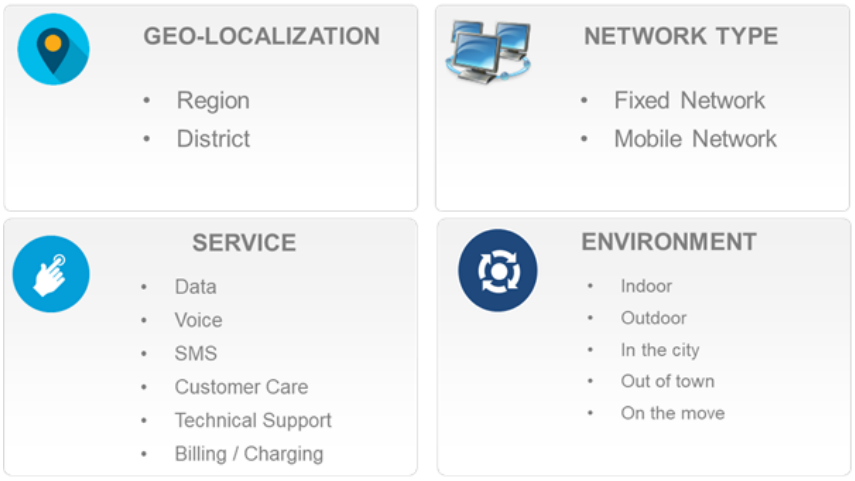
7. Public APIs
8. Open Data
9. Third party data
10. Own beacon data
11. Advertising data

At the beginning of the project, a subset of data sources has been selected (items #1, #2 and #3) as first sources of signal detection by crawling. However, while the analysis of the collected data moved forward, Twitter has being included among the monitored web data sources improving our understanding of the strength and significance of the signal in particular on social networks, as shown in the picture below.

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

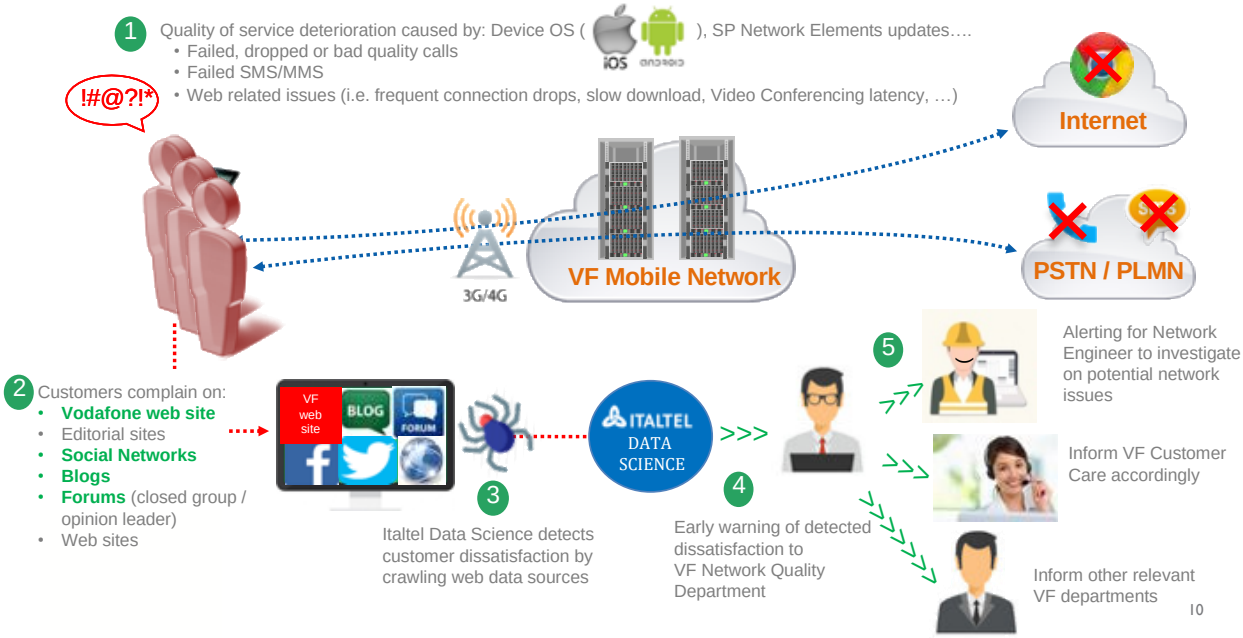


To signal qualification and source usability, the evaluation of the detected facts is added.
A constant statistical analysis highlights peaks of the dissatisfaction with a severity index and according to the following four analysis dimensions, where possible, as shown in the picture below.



Another important point is the evaluation of the minimum time frame for the availability of the new data flow, speed is a crucial factor as far as accuracy is concerned.
The following picture shows the high-level workflow of the use case.

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.



Early warning of detected complains to Vodafone Network Quality Department about a mounting customer dissatisfaction has beneficial impacts on:

Marketing

Customer experience improvement leads to both reducing the churn risk and improve the Net Promoter Score (NPS).

Customer Care

When the customer calls, a more informed call center knows the symptoms, most probable cause and resolution status, in order to deliver a more satisfying care experience, with higher first call resolution, shorter call handling time and reduced amount of calls. This definitely improves the customer care KPIs.

Operations

A faster identification and resolution of faults that are not detected during the standard network monitoring. As a consequence, Vodafone is able to deliver a better service with an enhanced customer satisfaction.

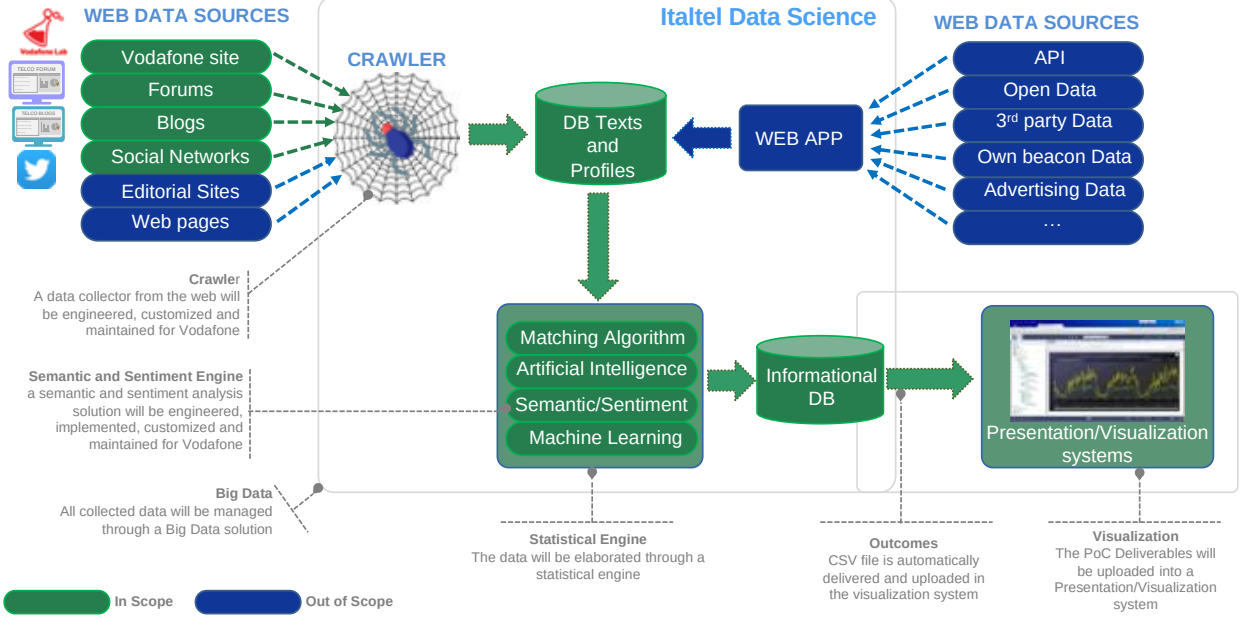
Planning

Prioritize network investment decisions based on knowledge of which upgrades will most impact the perceived customer experience.

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

Technical Solution

The picture below depicts the reference infrastructure of the proof of concept.



The solution is made up of three different components contributing with their characteristics and peculiarities in building a unique solution in the market:

1. Crawler
2. Semantic Engine
3. Statistical Engine

Crawler

A cutting-edge engine for the collection of data, developed since 2012 and based on an innovative concept: a bot that, automatically web scraping web pages like a human being, gathers the information from the visited web pages and structures them in relational database rather than a list of indexed pages. The engine is able to:

- Accurately visualize, read and save all the contents generated by means of browsing events (mouseover, ajax contents, …)
- Identify each area of the page (single articles, post, banner, …)
- Identify and click on links with the aim to display following contents (in case of gallery, in case of paging contents, …) keeping the relation with the original post
- Identify the elements inside the texts (single elements into the signature, link to social networks, recognize the original message from the quoted message)
- Follow the links keeping them related to the single text or user (links for further information, links to pages with its own profile, links to profiles of other forums, links to profiles inside social networks

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

- Interact with the APIs of any web site to integrate the data collected from the pages with the data available from the site itself
- Gather meta-information from the pages (metatag, RSS, configuration files, …) and link them with the data collected via scraping.

According to our goals, this approach, that maximizes and structures the information during the scan of the web (different from the classic Google method that indexes the pages to provide them as result of the searching engine), has many benefits:

- Identification of the single post of the single user (often included in a page with tens of posts)
- Connection with the topic of the thread or sub-thread (e.g. User post: *"It has not been working anything for three hours"* while the topic is "Issues in Tuscany")
- Connection with what it has been said in previous posts (e.g. User post: *"confirm"* but he/she is answering to a post that says: *"At Rome, 4G network does not work"*)
- Connection with what the user says in other posts (e.g. User post: *"Calls don't work also where I am"* and in other different post, maybe in the past, he/she said the he/she lived in Rimini)
- Creation of a single user profile (anonymized in irreversible way) very detailed and refined
- Connect with what the user says in his/her own profile (e.g. where he/she lives, which technologies is using, …)
- Connection with what the user states in his/her own signature (e.g. where he/she lives, which technologies is using, …)
- Connection with the user information collected outside the site under investigation (e.g. in other forums or on user's social pages).

Thereby, the proposed crawler has a capability to gather accurate information higher than any other crawler in the market.

Semantic Engine

The proposed sematic engine is more properly an ecosystem of different engines:

- NLP-Engine: A customized Natural Language Processing system able to identify short sentences with a great number of typical allusions
- Inheritance Engine: an engine able to understand the meaning of the messages analyzing the whole thread, not only the single message
- Stem and Lemmas Engine: engine able to identify the subject of every single term bringing it back to the unique lemma
- Entities Engine: engine that updates the typical technical terms of the context by crawling some reference sites
- Names Engine: engine that updates the typical proper nouns of the context by crawling some reference sites (data base of the devices, offers name, …)
- Derogatory Engine: engine that updates the database of the derogatory terms used to call the brand
- Taxonomy Engine: engine that uses a tree-lined taxonomy of concepts identified by keywords
- Meaning Engine: engine that understands: irony, double denials, implicit meanings

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

- Sentiment Analysis: engine that understands the sentiment of the sentence (positive, negative, sad, aggressive…).

Statistics Engine

Quantitative methods in combination with data science methods have been applied for the PoC Use case blending different engines that complete the solution and allow a numerical output of the results.

- **Supervised machine learning**

the engine is a neural network constantly educated by feedbacks from trainers who are responsible for correcting the erroneous analysis chosen on a sample from those with the lowest reliability index.

- **Items' weight engine**

A different weight is given to the items associated to the post (e.g. subject, location, type of problem, environment, device etc.) by inheriting them from other objects (e.g. title, previous posts in the thread, previous posts of the user, information retrieved from Social ...), in order to get a statistically robust allocation of these items to the post in question.

- **New entities evaluation engine**

This engine evaluates unknown terms with respect to typing errors, it is based on statistical models of frequency and type-error integrated with Apache Lucene fuzzy query.

- **Artificial Intelligence engine**

AI engine allows to evaluate the weight of heterogeneous sources based on various dimensions (frequency of posts, number of different people involved, rating, customized impact factor, number of citations ...). Several automatic processes continuously collect data from the web and a model based on Poset math models create a rank for every source.

- **Articles' weight engine**

This engine gives each post a weight taking into account different components (e.g. the robustness of the items, if data come from the sentiment, the site where it is posted ...)

- **Score engine**

This engine assigns a score to each analyzed dimension while maintaining the significant digits and which provides the final output.

Use Case Output

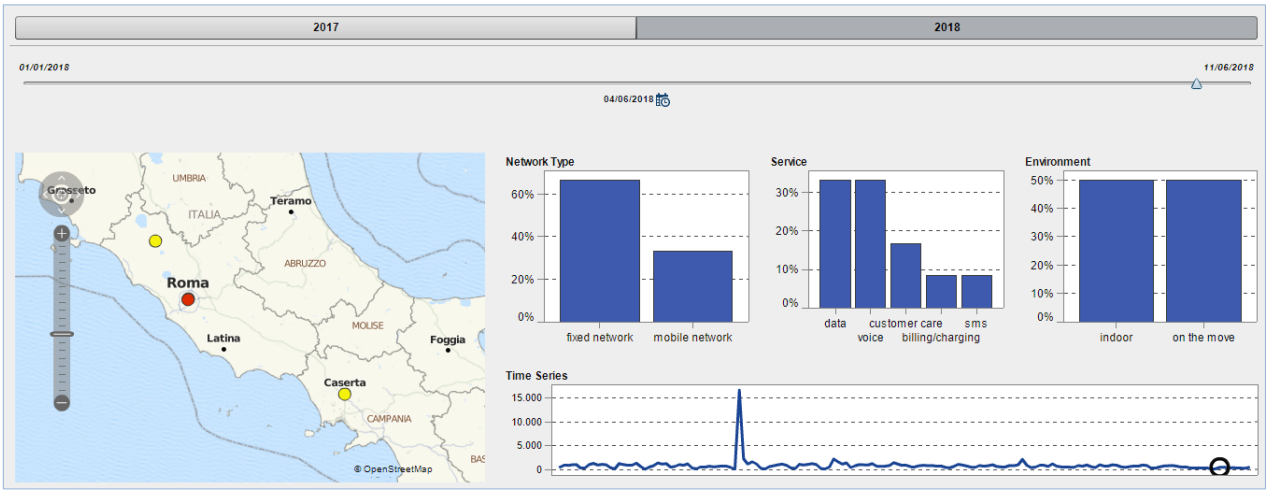
Expected output of the use case implemented are hourly csv files which indicate all the issues detected with their analysis dimensions and gravity score. The files can be imported in any analytics or alerting system. For PoC implementation, the output is gathered and visualized in a data presentation layer.

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

Data Visualization

An easy-to-use data presentation layer is necessary to quickly access, explore and analyze the produced data. The visualization acts in such a way to embody the data set and generate questions around it. Critical analysis of visualized data has been widely done as an important quality control on the collected data set, in addition to mere representation.
Geographic maps represent the complains distribution at a glance. They allow to highlight where complaints are located, get quickly insight of most critical areas and identify location-specific issues complains.
Interactive reports and dashboards include a variety of graphic objects (bar charts, time series), automatic filtering and facilities for data exploration.
Pre-defined reports provide several hourly, daily, weekly and monthly aggregations.

Below is an example of a daily view, the dashboard shows at a glance the four analysis dimensions and the time series of the selected year.



Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

Immediate Applicability

Italtel Data Science solution is applicable to the following use cases:

- **Early interception of Customers' dissatisfaction**

Using the hourly reports, it is possible to intercept the wave of discontent in the bud and before it reaches the maximum extent on Customer Care.

- **Maps overlap**

Overlapping the map (geographical but also service) detected by web analysis with maps of Service Provider's monitoring systems is possible:

  - identify areas where, although aware of having a less performing service, customers complain little, or vice versa, to detect areas where, while offering a good level of service, complaints are frequent. In this way, network upgrades can be planned based on customer satisfaction criteria
  - measure how much an encountered problem is actually perceived by customers and therefore be able to define the urgency of the intervention
  - evaluate the effects of network resources optimization activities
  - identify problems not detected by the monitoring systems and provide them with a new feed including new parameters to be measured or new kpi to be checked

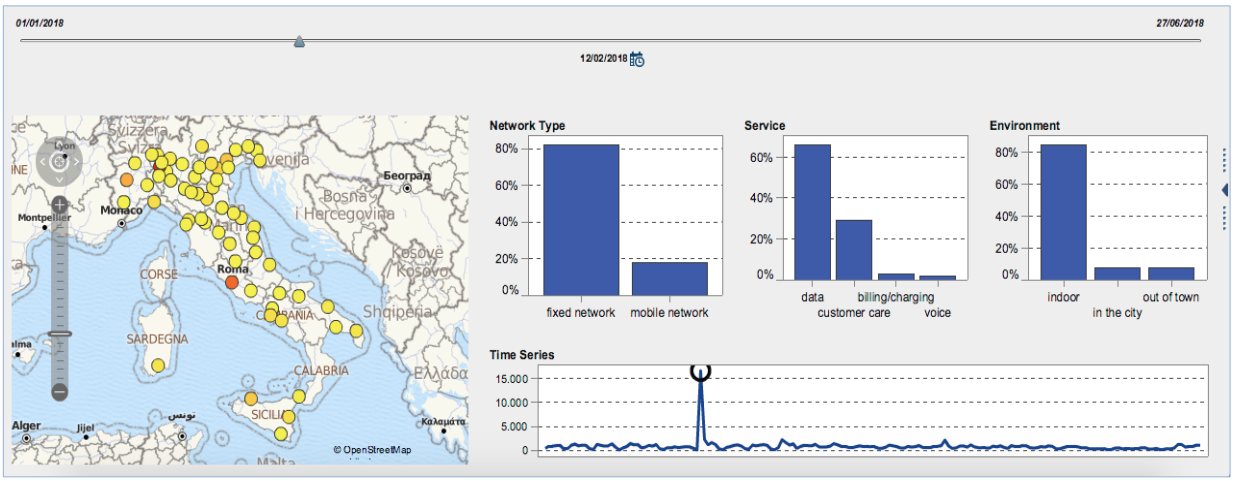
- **Data Correlation**

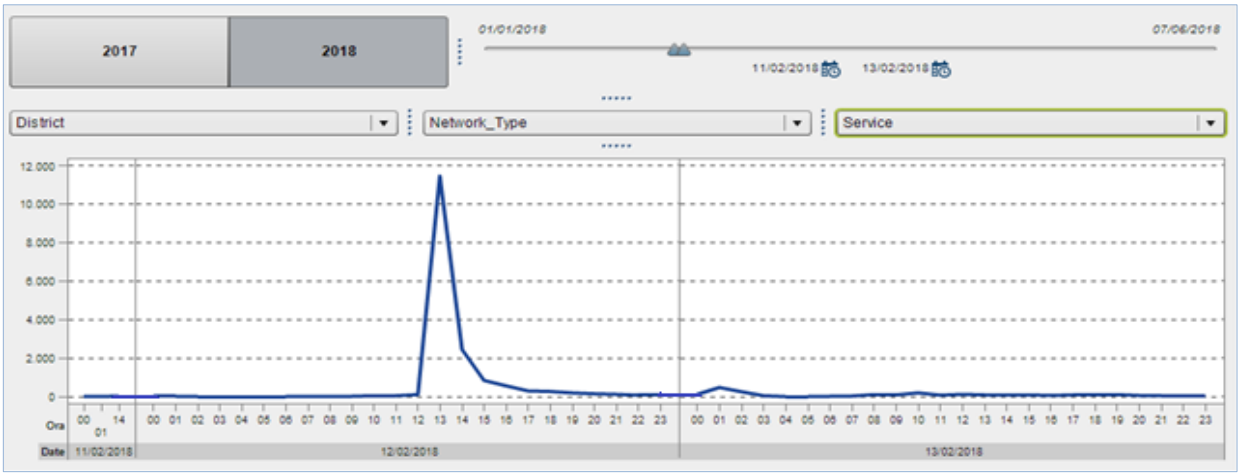
Data correlation with Network Data confirms the early detection of incidents. We made tests on issues recorded by Vodafone and our solution could detect 90% of the total significant issues that become 100%, if we exclude the rural areas, late night events or other situations in which typically people do not write on Internet.

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

Some Results

- **February 12, 2018**: Big problem in the Vodafone network throughout Italy. The system detected the problem between 12:00 and 13:00 am, while Vodafone detected the problem at 14:58, as reported in the first official tweet and shown at left-hand side below.

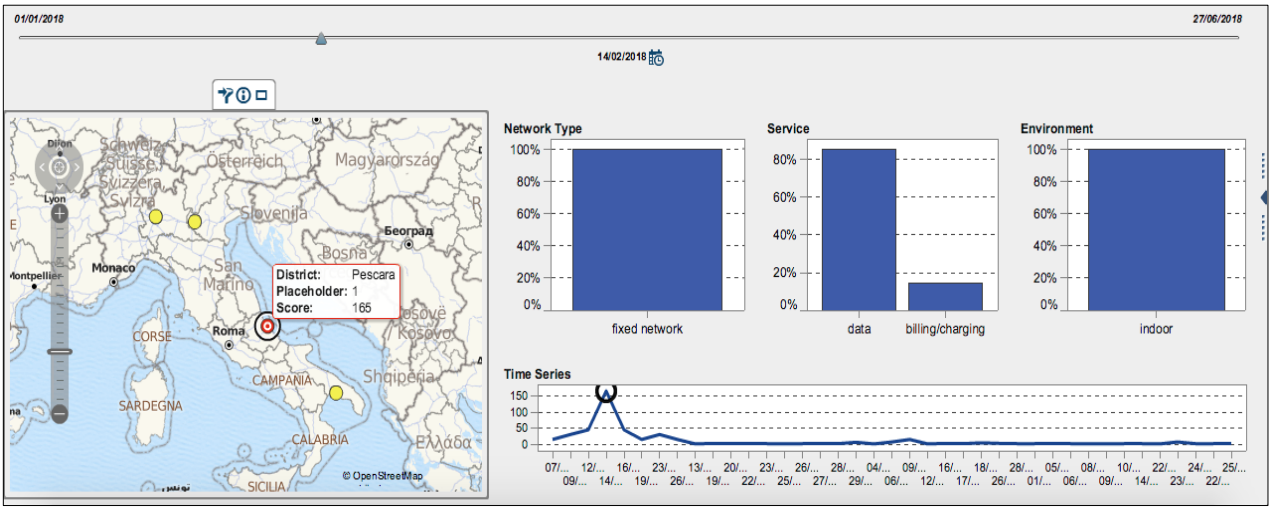


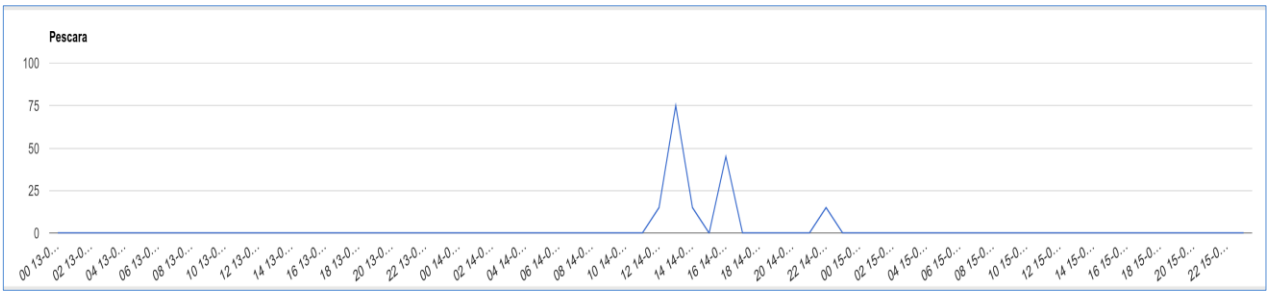




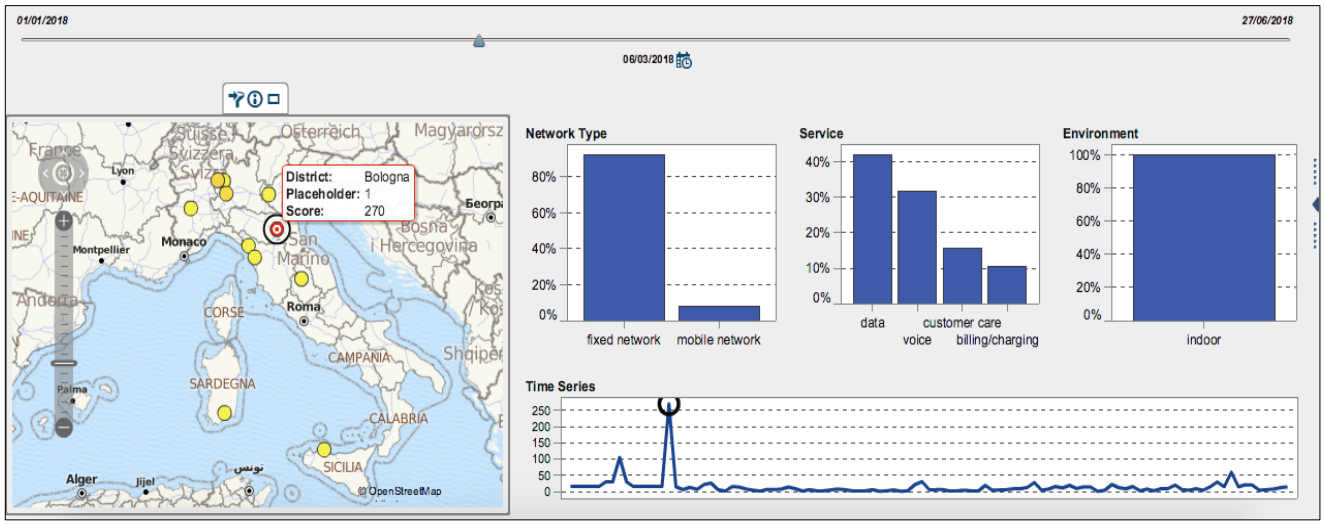
Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

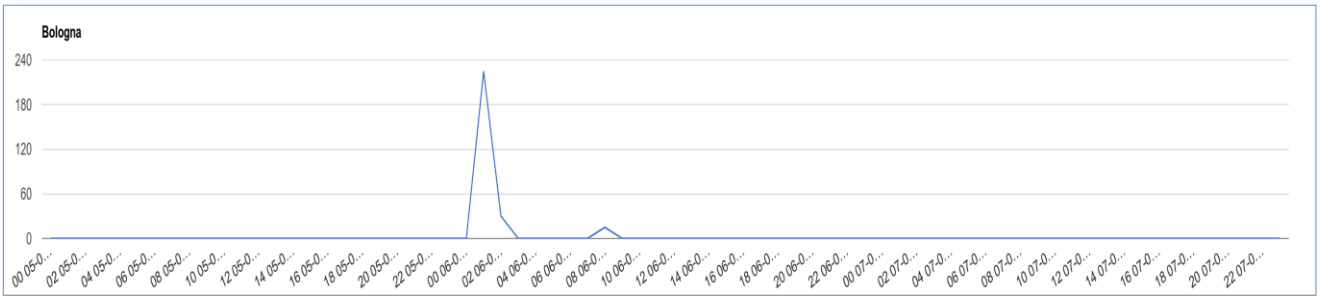
- **February 14, 2018:** Important issue on Pescara fixed network. Our system began to detect complaints between 11:00 and 12:00, while even the monitoring systems detected the problem at 12.30, we do not know when Vodafone became aware of the problem, demonstrating how often the issues are anticipated by malfunctions detectable from the web.





- **March 06, 2018:** Important issue detected on Bologna Fixed Network between 00:00 and 01:00 am, not detected by Vodafone.





Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

PoC evolution and other possible use cases

Key example applications include the following:

1. ***Evolution of Early Detection of Customer's Dissatisfaction Use case**:*
    a. **Enhance social networks analysis:** reinforce the signal adding other data sources like innovative digital customer care (e.g. *Facebook*..) and influencers blogs
    b. GUI evolution to represent additional details of the complains (e.g. specific keywords for each dimension, such as "*Android*", "*Vodafone station*", "*Fiber*"..)
    c. **Integration with SP's internal data, processes and systems** (e.g. Adding Trouble tickets source, converted by semantic analysis and then correlated to WEB data)
    d. **Applicability to other services**, enlarging the solution(e.g. AWS offering..)
    e. Applicability for other OpCos and networks contexts (e.g. roaming services..)
    f. **Listen to Customer real voice:** using speech to text capabilities, sentiment analysis and then analyzing the texts semantically
    g. Trend analysis on historical data
    h. Analysis of the seasonal nature of the problems
    i. Impact analysis of external events (matches, weather conditions..).
2. ***Competition Analysis**:* the same use case used for proprietary brand analysis can also be used to understand strengths and weaknesses of competitors, this allows to address specific marketing or commercial actions.
3. ***Other possible Data Science use cases:***

    Weighted Sentiment & CRM enrichment
    Churn Management
    Up/Cross-Selling
    Actual SIM users Identification
    Parental Control
    Real Time advertising optimization and Marketing Support
    Recommendation Engine

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

Components Proven Evidences

Crawler

Thanks to its great power and adaptability, the proposed crawler has been largely used in other contexts different from the one described in this whitepaper, such as:

- Continuous report of the editorials that refer to specific brand, company, product, …
- Reports on investments in advertising made both by a company for its own brand (investment campaigns check) and by competitors (benchmark)
- Editorials visibility index measurement (segmented by country, editorial groups, target, markets, other) and following the tracking history on different web pages and the related positioning
- Advertising visibility index measurement (segmented by country, editorial groups, target, markets, other)
- Relationship between the editorials visibility and the investments in advertising on a company's brand compared to reference competitors
- Local advertisement observatory (geo-localized and segmented by targets)
- National advertising observatory

Semantic Engine

Extensive experience in the implementation and use of semantic components within complex projects has been done in recent years:

- Online advertisement analysis
- Anti-churn solution
- User profiling by interest for CRM enrichment
- Market evolution analysis based upon consumer interests
- Trend analysis about politics subjects derived from institutional social networks

Statistic Engine

Our statistical engines analyze complex data and extract useful and effective information. They have been used in many real contexts and always with great success:

- Revenue forecasting (4% average error on the 14th day of the future)
- Optimization of the web advertising target (+600%)
- Ranking solutions with innovative models (500% more accurate than previous ones)
- Recommendation engine (600% click rate increase)
- Anti-churn prediction, adding a predictive evaluation to the first signs of churn intention, with the first ventile containing 400% of the churners compared to the last.

Final thoughts

This solution was a challenge faced together with Vodafone and based on the intuition that the web is a precious mine of information complementary to those that are usually collected by traditional systems and through probes and sensors. None of us knew if there would be a signal in the background noise of what customers write on the web.

The results show how the challenge has been won. This is really encouraging in view of the fact that this signal can only improve over time:

- the system will become more and more precise
- more and more people write on the WEB
- you can add other data sources that will improve information power and accuracy.

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.

Italtel is a multinational company that operates in the Information & Communication Technology sector with a strong focus on innovation. The Italtel offer is based on solutions for networks, data centers, business collaboration, cyber security, the internet of things. The solutions are composed of proprietary and third party products, managed services, engineering and consultancy services. Together with Exprivia, which recently acquired Italtel's controlling shareholding, it constitutes a new large Italian ICT group that intends to be a leader in digital transformation through the development of innovative solutions and services both in Italy and abroad. The target market consists of service providers, public and private enterprises, with a specific focus on vertical markets such as Energy, Health, Industry, Defense, Finance and Smart Cities.

Marketing_communication@italtel.com
www.italtel.com

Follow us on

Exprivia and Italtel integration creates a new industrial player
In Information & Communication Technology

Italtel and Italtel Logo are registered trademarks of Italtel SpA. All contents are Italtel SpA Copyright 2015. All rights reserved.
This document is intended for the addressee(s) only and is confidential and/or may contain legally privileged information;
it may not be reproduced or distributed in any form without Italtel's prior written consent.