Collaboration in Higher Education for Digital Transformation in European Business

The Age of Big Data Analytics: How Big Data is Reshaping the Insurance Industry

Margareta Teodorescu, margareta.teodorescu@fh-bielefeld.de
University of Applied Sciences Bielefeld/Germany;

Jan Budík, budikjan@gmail.com,
Brno University of Technology/Czech Republic

This working paper is a result of the EU Erasmus Plus project "Collaboration in Higher Education for Digital Transformation of Corporate Businesses" (CHEDTEB) and essential ideas come from the intensive exchange with IT experts and company managers of the working groups "Big Data" and "Blockchain technology". Further information about Big Data, algorithms, AI and Blockchain technology can be found on our project webpage.

http://www.chedteb.eu/

The authors are grateful for the lively discussions within various Big Data workshops and would like to thank in particular following colleagues for their valuable ideas, and suggestions:
Jan Luhan, Martin Fridrich – all of Brno University of Technology/Czech Republic,
Viire Täks. Luca Alfieri – both Tartu University/Estonia,
Rainer Lenz, Tahir Lushi - both University of Applied Sciences Bielefeld/Germany

Co-funded by the Erasmus+ Programme of the European Union

Obsah

I. Introduction

It is now uncontroversial that big data has become a real phenomenon, with a big impact upon businesses, individuals and the economy as a whole. Reports estimated a couple of years ago that by exploiting big data analytics at scale across their organizations retailers would increase their operating margins by more than 50 percent and that the healthcare sector would reduce costs by more than 5 percent[1]. However, the impact of data abundance extends well beyond the corporate environment.

Computer-automated analysis of blog postings, speeches, press releases or news articles, for instance, is also significant for the political environment, too. And this story repeats itself in sectors as varied as sports and science, public health and banking, agriculture and chemistry, marketing and advertising, and many more. In all these sectors there is a clear drive towards data-driven insights, discovery and decision-making.

It is no wonder then that these developments are now considered to be revolutionary: welcome to the Age of Big Data! A report by the World Economic Forum[2] declared data as being a new class of economic asset, similar to currency and gold. According to Gary King, director of Harvard's Institute of Quantitative Social Science,

> *"We're really just getting under way. But the march of quantification, made possible by enormous new sources of data, will sweep through academia, business and government. There is no area that is going to be untouched."*

One of the industries that is currently experiencing one of the major transformations of the past decades – the digital transformation – and which will be highly impacted by the Big Data phenomenon is, without question, the insurance sector. Insurance companies have to collect, store and evaluate large and diverse amounts of information and data.

As such, big data is especially promising for insurance companies and the use of big data can aid insurers in multiple ways, starting from underwriting and continuing with rating, marketing, and claim settlement practices. Data analytics in the insurance industry is transforming the way insurance businesses operate.

We propose in this paper an in-depth analysis of the telematics technology, a big data use case that provides solutions to car insurance challenges by reconfiguring risk-pricing in car insurance, enhancing insights for insurers and lowering premiums for responsible drivers.

---

[1] McKinsey Global Institute, *Big data: The next frontier for innovation, competition, and productivity* (2011)

[2] World Economic Forum, *Big Data, Big Impact: New Possibilities for International Development* (2012)

After introducing the new Pay-as-you-Drive (PAYD) approach in the motor insurance industry in the first chapter, the second one presents the four concepts for the technological implementation of PAYD in the insurance practice. The following three chapters focus on the GPS-based solutions and telematics (GPS-based PAYD-systems), analyzing the concept both from a business (Chapter III) and from an IT perspective (Chapter IV and V) as well. The last chapter concludes and summarizes the findings.

II. Pay-as-you-Drive – a new approach in the motor insurance industry

The term "Pay-as-you-Drive" (PAYD) refers to a new class of motor insurance, whose basic model of calculating premiums is no longer limited to individual and risk group specific features such as age, gender, vehicle and place of residence. The tariffing of this new class of car insurance includes, for the first time, also driving performance and driving behavior.

The use of new, modern technologies makes possible the reliable measurement and recording of these factors and allows the insurance companies to provide their customers with a broad range of innovative services. In the meantime, some insurance companies within the German-speaking area have initiated and launched several PAYD pilot projects.

The market for motor insurance is considered to be largely saturated and, based on demographic arguments, expected to shrink in the long run. In an era of ferocious competition, it is no wonder then that insurance companies are increasingly searching for new possibilities for differentiation and customer acquisition as well. Despite compulsory insurance in Europe and the US, motor vehicle insurance is only low margins.

PAYD insurance products can help differentiating under the pressure of competitors, addressing new customer groups and preventing or mitigating damages. PAYD is therefore not only a technical insurance concept, but also an important strategic option in the insurance industry, triggering low-cost and technology-driven motor vehicle insurance products.

In addition to these benefits for the insurance industry, PAYD concepts offer social benefits as well. According to a study by the Victoria Transport Policy (VTPI)[3], PAYD insurance is significantly more in line with the principle of causation. In contrast to conventional insurance products, which have to date, over-emphasized frequent travelers to the detriment of the other, non-regular travelers, PAYD systems address this disadvantage and thus contribute to fairer insurance products.

[3] Litman, T., *Distance-Based Vehicle Insurance Feasibility, Costs and Benefits. Comprehensive Technical Report*, Victoria Transport Policy Institute (2007)

The insurance practice has developed four concepts for the technological implementation of PAYD. They differ significantly from each other in terms of technical complexity, actuarial accuracy as well as social and economic feasibility.

1) Mileage Rate Factor

This concept is widely used by numerous insurance companies and does not require any additional technology. In this case, the policyholder estimates his annual mileage, indicates this at the conclusion of the contract and adapts it by contract renewal or anytime when living conditions have changed. The policy consists of a fixed part covering the general risk characteristics such as age and vehicle features, and a variable part that considers the mileage. Even though this concept is well established and widespread, it has important drawbacks.

2) Usage-Based Premiums

This second concept extends the principle of Mileage Rate Factor. Instead of just requiring estimates of annual mileage, the insured regularly reads and reports the mileage to the insurer. This can be done by the customer himself or during the annual inspection. An automatic, technology-based monthly transmission of the mileage to the insurance company is also possible. So far, the Japanese Aioi Insurance Company has successfully implemented this solution.

3) Pay-at-the-Pump

The Pay-at-the-Pump concept assumes that the insurance premium is directly and already included in the fuel price and, as such, it replaces the regular insurance payments. Even though many experts in the past recommended this concept, it has been introduced solely in South Africa, and covers just a minimal insurance for traffic accidents.

4) GPS-based solutions

The most up-to-date and technically sophisticated concept is GPS-based PAYD solutions. Not only do these solutions make possible to determine the number of kilometers driven, but they also record all details about when, where and how the vehicle is used, in order to calculate the individual insurance premium. This last concept will be detailed in the next chapter.

GPS-based solutions and telematics

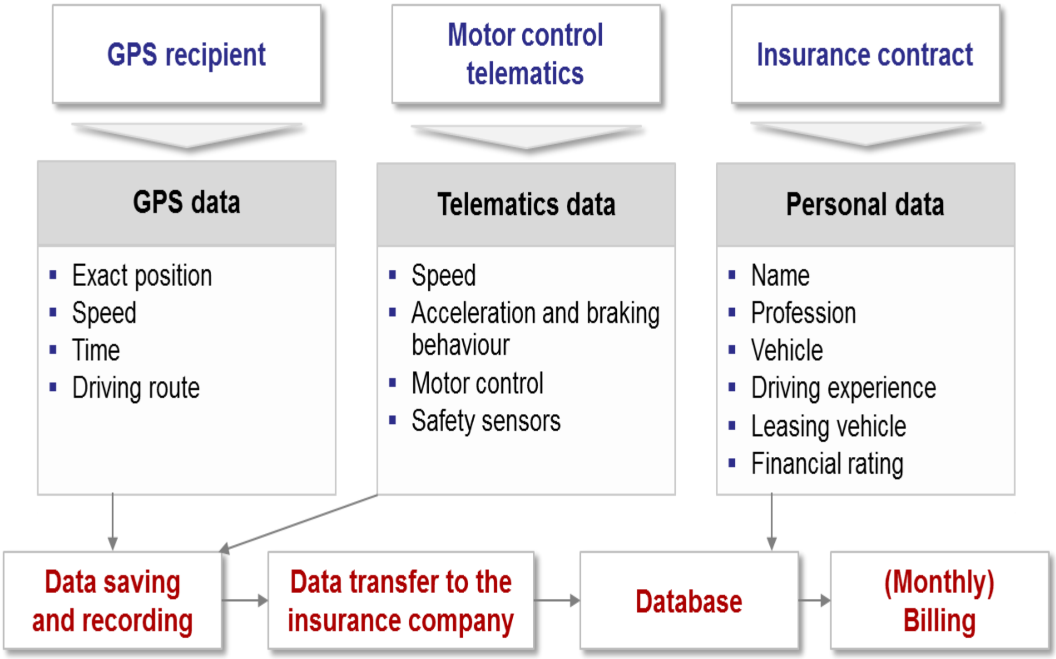
The word "telematics" stems from the combination of telecommunications and informatics, and it was the joining of these two sciences that resulted in the field of telematics. In its broadest sense, telematics is based on the Internet itself, as it basically combines telecommunications of any kind (phone lines, cables, etc.) with informatics, such as computer systems. The term is used most commonly in the automotive and, respectively, insurance industry and refers to vehicle telematics, where location information is used in different insurance business applications.

The British insurance company Norwich Union has registered one of the most consistent implementation of this PAYD concept, which uses a large body of data for tariff classification and security-related services. For young drivers, for instance, the company offers a policy that calculates the kilometer-based tariff separately for two periods. In the normal period (06:00 – 22:59), the premium per kilometer is calculated according to the personal risk profile of the insured and costs only a few British pence. During the risk period (23:00 – 05:59) the insurance rate is £1 for each completed kilometer. This is intended to encourage novice drivers not to drive during the night, when the risk of traffic accidents is highest.

The Royal & Sun Alliance has launched another simple, GPS-based PAYD model for novice drivers. Also this model distinguishes between driving during a risky and a less-risky period of time, but only the starting and ending times of the journey are reported to the insurer. In case that the drive takes place during the risky time between 23:00 – 6:00, the driver will then be charged a flat rate of £ 25.

The GPS-based PAYD system offers an important financial incentive with respect to the number of completed kilometers. The incentive tends to be significantly higher, the greater the reduction in mileage for PAYD customers. Since this is a kilometer-based PAYD solution, it is a worthwhile alternative primarily for conventional drivers who are not frequent drivers. Therefore, the reduction of mileage and the associated pollutant emissions can be much lower than expected.

Figure 1: The operating mode of GPS-based PAYD-systems[4]



The technical implementation of PAYD-systems provides the insurance companies with a relatively high freedom of design. The technology used is oriented and adapted to the operation scope and planned functionality of the PAYD product. The implementation of a GPS-based PAYD concept is currently realized within most insurance companies by the installation of appropriate telematics hardware. This typically has to fulfill four tasks.

a) Capture the GPS signal

Capturing the GPS signal requires a suitable receiver, which has to be placed as close as possible to the windshield of a vehicle, in order to ensure the best possible reception. If a car is already at the factory equipped with a GPS receiver, then the installation of an additional receiver is obviated. If, otherwise, the distance covered is not calculated via GPS and no other information about the type of roads used is required, then the installation of a GPS module is not mandatory. In this case, the mileage will be recorded and read together with other telematics data directly from the on-board electrical system.

[4] Based on (i.a.) Ippisch, T., Thiesse, F., *Das Pay-as-you-drive (PAYD)-Konzept in der Versicherungswirtschaft* (2007)

b) Record the telematics data of the vehicle

Numerous different telematics data in modern vehicles today are being generated and continuously evaluated by the on-board computer. Depending on how the PAYD concept has been designed, data such as mileage or speed can be read from the corresponding telematics modules. This can be further extended to include capturing accelerometer or airbag sensor data.

c) Storage of the data

The storage of GPS and telematics data usually takes place directly on the telematics modules. A first evaluation of the data is usually made, such that only metadata can be stored. In extreme cases, the complete premium calculation can be done onboard and, subsequently, the monthly premium amount is reported to the insurer.

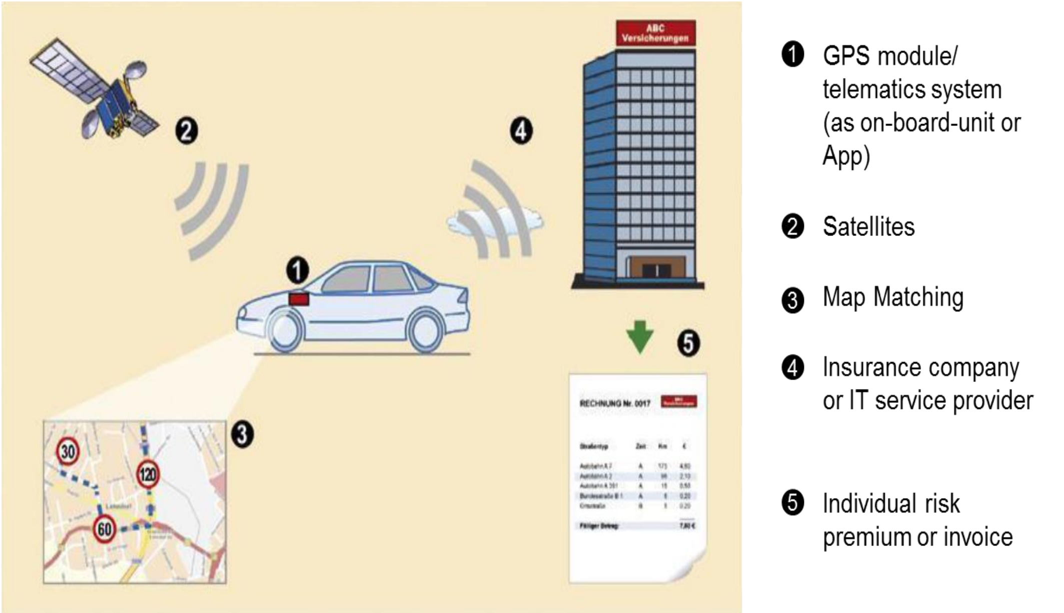
d) Report the data to the insurance company

While capturing the GPS and telematics data as well as their storage only slight differences in technological implementation need occur; during the final transmission of the data different approaches can be pursued. The most customer friendly is an automatic, regular transmission of data via GPS. Norwich Union uses this method, for example, for its PAYD products.

Despite being very comfortable for the customer, this solution exhibits an important cost-related disadvantage. It is not only a GPRS-enabled telematics module being necessary, but also a mobile phone contract, which involves additional fixed and variable costs. The regular data transmission alone leads to estimated monthly additional costs of about € 10.

The GPS-based PAYD solution is being viewed critically also for privacy reasons, as it is feared that by combining GPS and GPRS the location of a person could be determined and reported at any time. So far, however, there is no PAYD concept providing continuous transmission of GPS location information. As such, this concern appears to be unfounded. Only fleet management systems such as NavTrack, TomTom Work and FleetBoard allow the determination of vehicle positions in real time.

Figure 2: An illustration of the operating mode of GPS-based PAYD-systems[5]



Regardless of which technology is used for data transmission, all information is transmitted to the telematics server of the insurance company. In case that GPS position data is also transmitted, this is matched on the telematics server with the corresponding maps. In this way, the number of traveled kilometers and the different road types as well can be determined. Based on this information and on the customer information that is being deposited on the server, the insurance company can regularly calculate the individual insurance premium. Figure 2 provides an illustration of how the GPS-based PAYD-systems are operating.

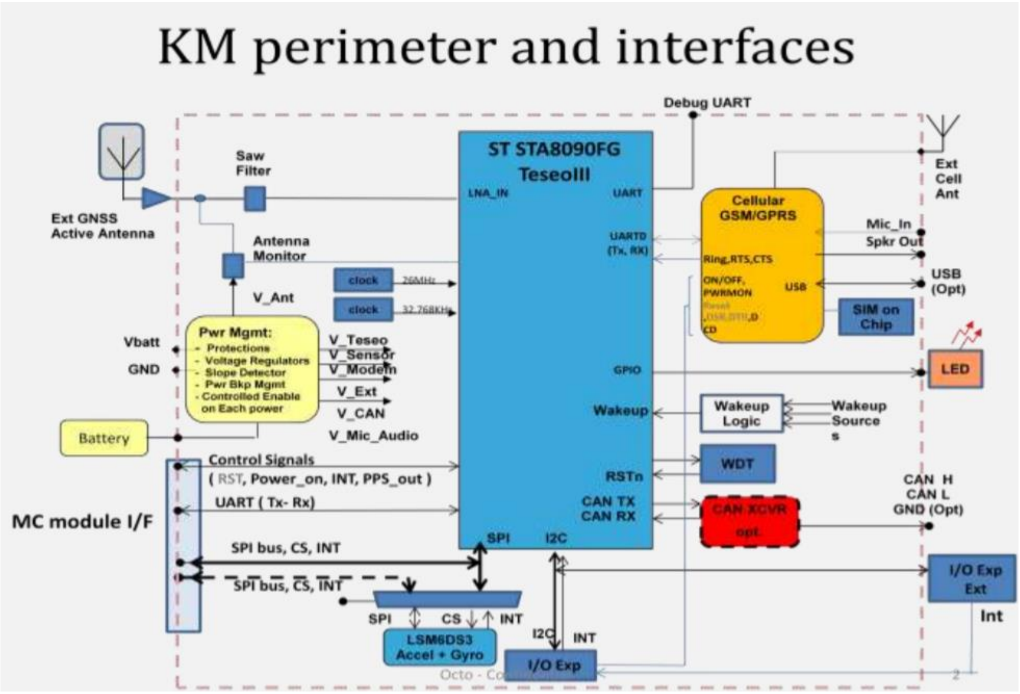
The telematics solution is based on a telematics control unit, which is an embedded hardware and software system. By wirelessly connecting a car – typically via a cellular modem – to its surroundings and beyond, the telematics control unit sources and/or exchanges information from/with other systems. The following two chapters will give an overview about the mostly used telematics hardware and software technologies, respectively.

[5] Lochmaier, L., *Pay-as-you-drive: Auto-Versicherer planen User-Tracking vis GPS*, in Bohman, R. (Hrsg.), *monitor – Das Magazin für Informationstechnologie*, p. 41 (2007)

III. Telematics hardware technologies

One of the most frequently used telematics hardware solution in the insurance industry is ZTE AT21, which contains all the hardware necessary for implementing the insurance telematics product. The first part of the insurance telematics technology is the hardware, which basically allows two processes: data measurement and data transfer. At a first glance it may seem very simple, but telematics hardware needs to secure precise communication between different sensors in real time. The architecture of the Kernel module of ZTE AT21 is depicted in the following figure.

Figure 3: The architecture of the Kernel module of ZTE AT21[6]



The main hardware architecture features and physical constraints are:

- GNSS multiconstellation
- GSM
- Accelerometer
- SIM on Chip
- Watch Dog
- Power Management
- RF transceiver
- CAN Bus interface

[6] Octo Users' manual, retrieved from https://usermanual.wiki/Octo-Telematics-S-P-A/ZTEAT21/html

GNSS multiconstellation is the ability of GNSS receiver to work with multiple frequencies from multiple constellations in position calculation, allowing for minimizing position errors. The multi-constellation receiver can access signals from GPS, GLONASS, BeiDou and Galileo as well.

This multiple connections technology definitely offers several advantages, like allowing for a reduced signal acquisition time, improved position and time accuracy, reduction of problems caused by obstructions such as buildings and foliage, and improved spatial distribution of visible satellites, resulting in improved dilution of precision[7].

The next main hardware architecture feature, GSM – shortcut for Global System for Mobile Communications –, is the most widely used wireless technology in the world, available in more than 200 countries with a market share of more than 90 percent. This technology allows the telematics hardware to send data directly to the insurance company[8].

The accelerometer is an electronic component that measures tilt and motion. It is also capable of detecting rotation and motion gestures, such as swinging or shaking. The most common use for it is to activate auto screen rotation on mobile devices when the user changes their orientation from portrait to landscape or vice-versa. The accelerometer has the function of a car acceleration measuring device and can detect car crash as well[9].

Another important element of the hardware architecture is the SIM (Subscriber Identification Module), which is a small card that identifies a mobile device on a cellular network. It contains an integrated circuit that stores a unique identifier called an "international mobile subscriber identity" (IMSI) number and other information specific to the mobile carrier. In the telematics unit this is directly implemented on chip[10].

A watchdog is a device used to protect a system from specific software or hardware failures that may cause the system to stop responding. The application is first registered with the watchdog device. Once the watchdog is running on the system, the application must periodically send information to the watchdog device. If the device doesn't receive this signal within the set period of time, it would execute the proper keystrokes to reboot the machine or restart the application[11].

---

[7] Source: https://www.novatel.com/an-introduction-to-gnss/chapter-5-resolving-errors/multi-constellation-and-multi-frequency/

[8] Source: https://web.archive.org/web/20140208025938/ http://www.4gamericas.org/index.cfm?fuseaction=page&sectionid=242

[9] Source: https://www.gsmarena.com/glossary.php3?term=accelerometer

[10] Source: https://techterms.com/definition/sim_card

[11] Source: https://www.webopedia.com/TERM/W/watchdog.html

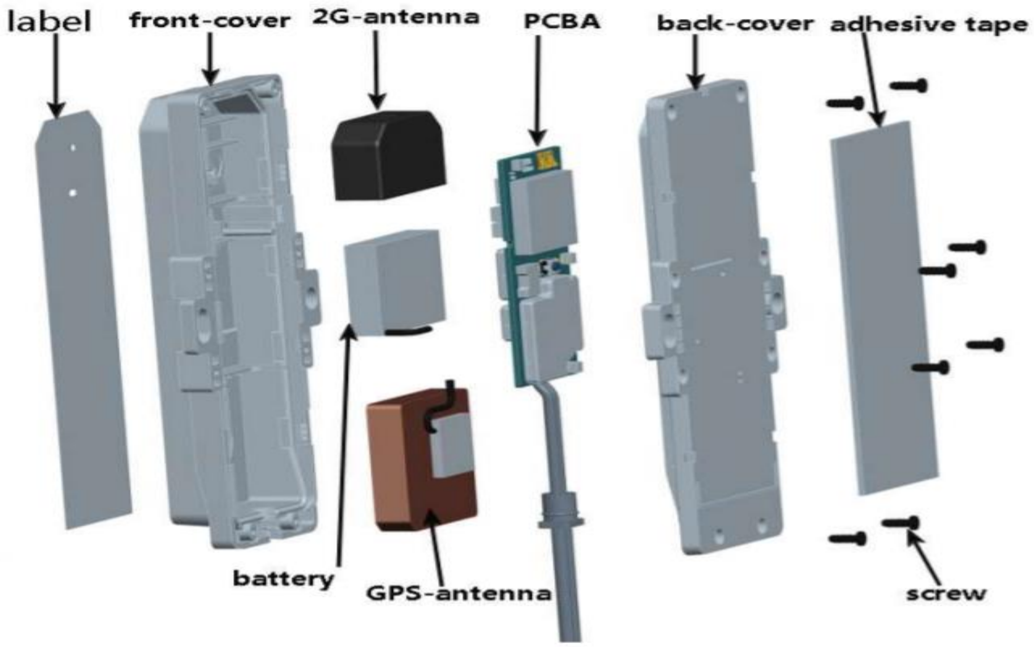
Power Management provides power protection to be compliant with E-mark, voltage regulators to feed different modules, slope voltage detectors, and voltage temperature measurements for battery back-up management[12].

Transceiver is the term used for devices which house both the transmitter and the receiver in a single module. Such devices which transmit and receive Radio Frequency (RF) signal are called RF Transceiver. The RF Transceiver position is mostly between Baseband MODEM and PA/LNA in any wireless communication system. RF Transceiver is used to convert IF frequency to RF frequency and vice versa. It is used in satellite communication, for radio transmission and reception, for television signal transmission and reception, and in wimax/wlan/zigbee/lte networks[13].

Although the CAN bus protocol was introduced for automotive use, the system is used today in several other networking applications. With the simple system of networking individual modules into a single central line of control, many companies are taking advantage of the CAN bus protocol in assembly lines, industrial vehicles, medical machinery and any other machinery that uses multiple modules that communicate with each other[14].

The following figure depicts some components of the telematics hardware.

Figure 4: Telematics hardware[15]



[12] Source: https://usermanual.wiki/Octo-Telematics-S-P-A/ZTEAT21/html
[13] Source: https://www.rfwireless-world.com/Terminology/rf-transceiver.htm
[14] Source: https://sewelldirect.com/learning-center/canbus-technology
[15] Telematics hardware, retrieved from https://apps.fcc.gov/eas/GetApplicationAttachment.html?calledFromFrame=Y&id=3108306

IV. Telematics software technologies

The telematic hardware ZTE AT21 sends large amounts of data about each user through the GSM network to the insurance company. These data contains information about the exact position, speed, time and all details about the driving route. For hundreds of thousands of users, the size of this data can be around multiple gigabytes per day. As such, the insurance companies need to solve three basic processes: data storage, customer classification and analysis, and visualization.

Data storage

The storage of increasing amounts of data require using new technologies. One of the best services for Big Data storage is Amazon Web Service (AWS). AWS offers on demand cloud computing platforms, which can be used by individuals, companies or governments and is based on a pay-as-you-go system. The leader in Big Data solutions is Apache Hadoop, which is an open source software project that can be used to efficiently to process large datasets. Instead of using one large computer to process and store the data, Hadoop allows clustering commodity hardware together to analyze massive data sets in parallel.

Hadoop is commonly used to process big data workloads because it is massively scalable. Hadoop provides a high level of durability and availability, while still being able to process computational analytical workloads in parallel. The combination of availability, durability, and scalability of processing makes Hadoop a natural fit for big data workloads[16].

Customer classification and analysis

The collected data is analyzed by the insurance companies, in order to get more information about each user and try to understand his/her behavior. For this reason it is necessary to use sophisticated methods like machine learning. One open source framework is Scikit-learn, which is designed for Python data scientists. This is a simple and efficient tool for data mining and data analysis, and is accessible to everybody. This framework can do the following:

- classification
- regression
- clustering
- dimensionality reduction
- model selection
- preprocessing

[16] Source: https://aws.amazon.com/emr/features/hadoop/

In case of an insurance company, this data mining framework can very quickly analyze each user in terms of speed, the most common time when a car is driving, and the exact location where the user commonly drives. The user can be classified, for example, in the following ways:

- most often user drives the car between 9am and 11am
- in 43% cases user drives on the highway
- in 70% cases user accelerated from 0 km/h to 50 km/h in 4.5 seconds
- user often overtakes other cars

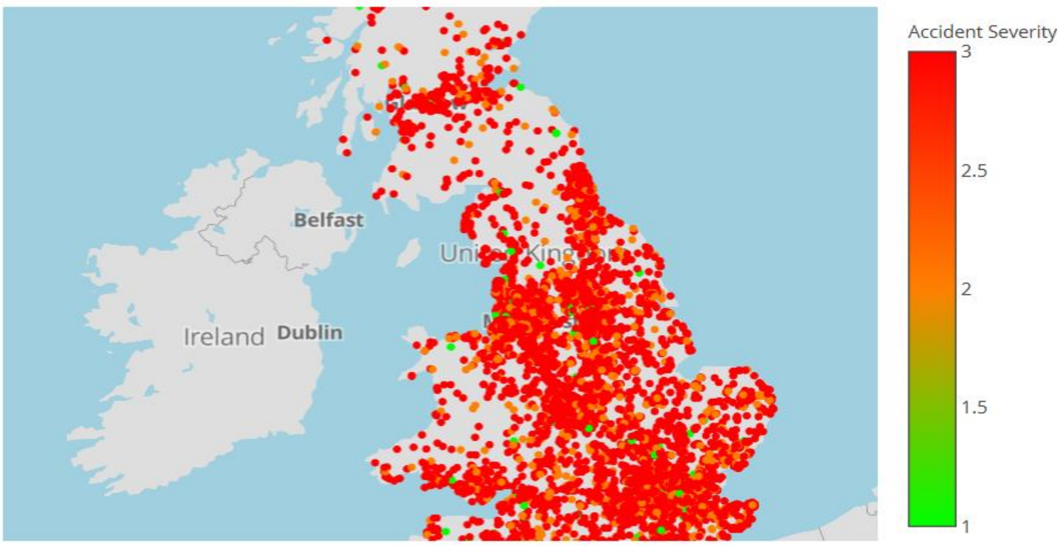
This type of classification can be used for insurance product personalization.

Visualization

The visualization problem can be solved by using a graphical library. If data scientists would like to just analyze graphically for example the density of car parking, traffic jam heat map or individual users' positions during the day, they need to use some graphical library. Python developers commonly use Plotly.

Plotly provides online graphing, analytics, and statistics tools for individuals and collaboration, as well as scientific graphing libraries for Python, R, MATLAB, Perl, Julia, Arduino, and REST. Based on this analysis, insurance companies can visualize for example the car accident situation and, accordingly, change the insurance agreement[17].

Figure 5: UK road accidents from 2012 to 2014[18]

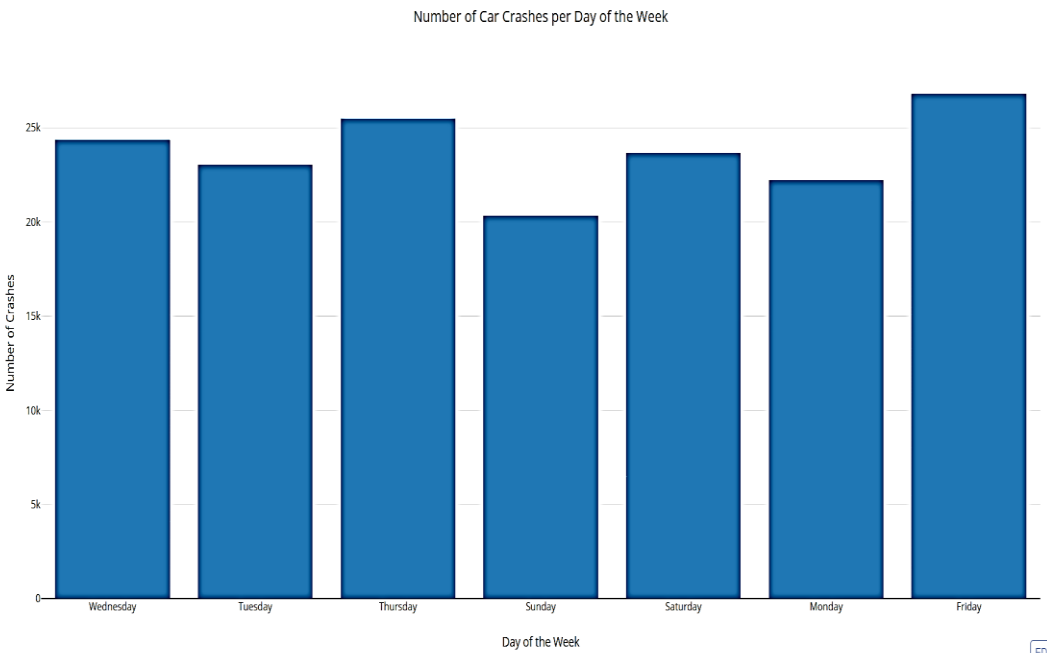


---

[17] Source: https://plot.ly/

[18] Source: UK road accidents, retrieved from https://plot.ly/~slavino95/5/uk-road-accidents-from-2012-to-2014/#/

Figure 6: Car crashes per day of week[19]



V. Concluding remarks

This paper provides an overview about the development of telematics and usage-based car insurance as a Pay-as-you-Drive insurance product, which has been identified as an opportunity to attract customers through customized premium models and benefits.

The paper began by introducing the new Pay-as-you-Drive (PAYD) approach in the motor insurance industry and by presenting the four technical implementations of the PAYD insurance model, namely Mileage Rate Factors, Usage-Based Premiums, Pay-at-the-Pump and, respectively, GPS-based solutions for PAYD models.

These four concepts differ significantly from each other in terms of not only technical complexity, but also with respect to actuarial accuracy and social and economic feasibility. The following three chapters present the GPS-based solutions for PAYD models by analyzing this concept from both a business and an IT (hardware as well as software) perspective.

[19] Source: Car crashes per day of week, retrieved from https://plot.ly/~cvanderw/81.embed

The development of telematics and usage-based car insurance began in the early 2000s and – due to the rise of mobile data, smartphones and, increasingly, connected cars – has now evolved to a mature level.

Telematics provides a full, complete picture of drivers' behavior, at any point in time and at any location. By effectively processing the data about when, where and how an insured person is driving, insurance companies can extract a very accurate assessment of individual driving behaviors and risk profiles. Based on this, insurers can derive the likelihood of that person causing an accident and, subsequently, making a claim[20].

Linking the information between premium pricing and driving behavior, telematics brings clarity and customization to both insurers and drivers. Furthermore, recent development of software and mobile apps have enhanced the possibility for drivers getting feedback on their driving performance. Based on this feedback, drivers are encouraged to adjust and adapt the way they drive and, as such, to take a more active role in influencing and determining the car insurance premium they pay.

By enhancing transparency around premiums, telematics has succeeded in revolutionizing premium pricing. Big data has become by now an all-encompassing sphere, where the existing digital, analytical and statistical capabilities make use of traditional sets of data in order to analyze drivers' behavior and extract information from it that makes possible customization of insurance products.

[20] Patrick Quinn, *Opening the black box on premiums: three ways that big data is changing car insurance models*, retrieved from https://www.insurancebusinessmag.com/uk/opinion/opening-the-black-box-on-premiums-three-ways-that-big-data-is-changing-car-insurance-models-111423.aspx

References

Huang, C.-M., Chen, Y. S., *Telematics Communication Technologies and Vehicular Networks: Wireless Architectures and Applications*, Information Science Publishing, IGI Global (2009)

Ippisch, T., Thiesse, F., *Das Pay-as-you-drive (PAYD)-Konzept in der Versicherungswirtschaft, Innovative Services und attraktive Prämienmodelle durch vermehrten Technologieeinsatz in der Versicherungsbranche*, M-Lab Arbeitsbericht 34, Universität St. Gallen (2007)

Karau, H., Konwinski, A., Wendell, P., Zaharia, M., *Learning Spark: Lightning-Fast Big Data Analysis*, O'Reilly Media, Inc. (2015)

Litman, T., *Distance-Based Vehicle Insurance Feasibility, Costs and Benefits. Comprehensive Technical Report*, Victoria Transport Policy Institute (2007)

Lochmaier, L., *Pay-as-you-drive: Auto-Versicherer planen User-Tracking vis GPS*, in Bohman, R. (Hrsg.), *monitor – Das Magazin für Informationstechnologie*, p. 41 (2007)

McKinsey Global Institute, *Big data: The next frontier for innovation, competition, and productivity* (2011)

Patrick Quinn, *Opening the black box on premiums: three ways that big data is changing car insurance models*, retrieved from https://www.insurancebusinessmag.com/uk/opinion/opening-the-black-box-on-premiums-three-ways-that-big-data-is-changing-car-insurance-models-111423.aspx

Tan, P.-N., Steinbach, M., M.,Karpatne, A., Kumar, V., *Introduction to Data Mining, What's New in Computer Science*, Pearson (2018)

World Economic Forum, *Big Data, Big Impact: New Possibilities for International Development* (2012)