Technological Forecasting & Social Change 170 (2021) 120865

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Door-to-door air travel: Exploring trends in corporate reports using text classification models

Ulrike Schmalz [*,a,b], Jürgen Ringbeck [b], Stefan Spinler [b]

[a] Bauhaus Luftfahrt e.V., Willy-Messerschmitt-Str. 1, 82024 Taufkirchen, Germany
[b] WHU - Otto-Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany

ARTICLE INFO

Keywords:
Annual reports
Door-to-door air travel
Small data
Support vector machine
Text classification,

ABSTRACT

Previous studies have identified key trends affecting the door-to-door air travel chain, to inform organizational decision making. However, the extent to which transport service providers consider strategically relevant trends remains unclear. This study adopts the novel scope of door-to-door air travel in applying multi-labeled text classification models to 52 corporate reports from a sample of transport service providers. Trends identified from a literature review are used to develop seven classes. Two prototype models are developed: a dictionary-based classifier, and a supervised learning model using the multinomial naive Bayes and linear support vector machine classifiers. The latter yields the best model output. The results reveal that providers consider environmentally-friendly air transport and related products to be highly relevant, while disruption management, leveraging passengers data and improving airport feeder traffic through novel mobility concepts are considered to be of medium relevance. These models enable cheaper and quicker analysis of companies textual data. This innovative approach is also applicable to other research questions, such as market studies and finance-related projects.

1. Introduction

In light of the COVID-19 pandemic and its fundamental impact on air travel, it may become more important than ever to offer bundled and contactless door-to-door (D2D) air travel to regain passengers trust and meet hygiene regulations. In pre-COVID times, air travelers complained (Schmitt and Gollnick, 2016) about factors such as lack of comfort (Baumgartner et al., 2016; Sezgen et al., 2019), non-transparent pricing (Budd et al., 2016), long travel times (Ureta et al., 2017) and disruptions (Kim and Park, 2016). Many of these pain points affect not just a particular flight segment but all parts of the journey. Exploring the entire D2D travel chain is indispensable to tackling these problems and improving the overall journey experience for air transport passengers. The scope of D2D is of high practical relevance to the industry, and had already been recognized pre-COVID by organizations such as Lufthansa (Lufthansa Innovation Hub, 2020c), IATA (IATA, 2020b) and Boston Consulting Group (Wade et al., 2020). D2D solutions may open up business opportunities for transport service providers (TSPs), such as additional revenues and market share from innovative and integrated travel products, and the potential to introduce novel airport access modes, such as urban air mobility, airport car sharing and high-speed transrapid systems. Environmental debates also influence air travel chains, giving rise to flight shaming, less willingness to fly and a concern to reduce carbon dioxide emissions.

Extant studies discuss factors affecting many parts of the D2D air-travel chain (Kluge et al., 2020). To remain competitive, the industry supply side (airlines, airports, airport feeder providers) must understand these factors and respond to novel developments (Wade et al., 2020). Some studies highlight their managerial implications, but none has so far examined whether TSPs act on them. This paper fills this gap by presenting an innovative approach to uncovering the relevance of strategic aspects of the supply side. We ask *to what extent do TSPs consider prevailing factors in their strategic planning and incorporate them into their communications?* To answer this question, we apply two novel text-classification models to textual organizational data drawn from providers covering the five main segments along the D2D air-travel value chain. We determine strategically relevant trends in the dataset that must be accounted for in organizational decision making. These aspects are represented by seven classes, developed from hypotheses based on a literature review of mobility research.

Text analytics, as part of machine learning, is already used in the industry. A recent survey of 295 German companies reveals that around

* Corresponding author.
E-mail addresses: ulrike.schmalz@bauhaus-luftfahrt.net (U. Schmalz), juergen.ringbeck@whu.edu (J. Ringbeck), stefan.spinler@whu.edu (S. Spinler).

https://doi.org/10.1016/j.techfore.2021.120865
Received 2 February 2021; Accepted 1 May 2021
Available online 23 May 2021

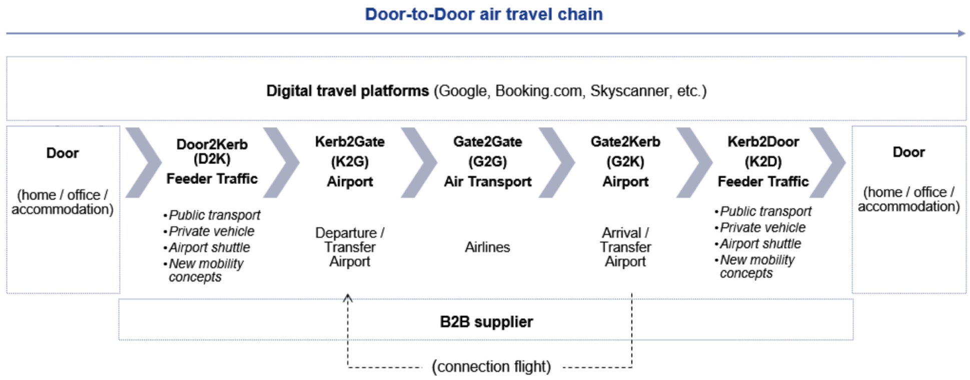

Fig. 1. Transport service providers along the D2D air-travel chain (authors depiction).

46% already apply text analytics (IDG Business Media, 2020). Among other metrics, the Lufthansa Innovation Hub uses simple keyword frequency counts from annual reports and social media platforms to develop the airlines digital index (Lufthansa Innovation Hub, 2019). Text analytics also has implications for finance and accounting (Loughran and Mcdonald, 2016). Bloomberg Professional Services uses natural language processing to retrieve valuable company information from unstructured text, such as news articles and social media data, applying the support vector machine (SVM) and the K-nearest neighbors (KNN) algorithm in its text classification model. Insights from the models results support investment decisions (Bloomberg Professional Services, 2018).

This paper is structured as follows. Section 2 presents a definition of and trends in D2D air travel, elaborates on previous text analytics studies and highlights this studys contributions. The methodology is described in Section 3. Section 4 explains the development of the prototype text classification models and their outcomes, which are discussed in depth in Section 5. Section 6 specifies some limitations of this study, suggests avenues for further research, and draws some conclusions.

2. Definition and previous work

After defining D2D air travel, this section reviews key trends in this area, presented in terms of seven trend hypotheses, and explores studies that have leveraged textual data.

2.1. Door-to-door air travel

D2D air travel is defined in this study as the physical movement of passengers from origin to final destination, including all modes of transport, transfer and supporting activities (see Fig. 1). The study focuses on travel chains that include air transport, so everyday mobility, vacation packages offered by travel agencies and hospitality services are not within its scope. It considers both passengers (demand side) and companies offering mobility products and services (supply side), also known as TSPs.

The D2D air-travel chain consists of five main segments[1]. In developed markets such as Europe, feeder traffic for airport access and egress is covered by public transport, private vehicles and individuals walking or cycling (Budd et al., 2016), as well as by new sharing mobility concepts such as Uber (Young and Farber, 2019). These door-to-kerb[2] (D2K) and kerb-to-door (K2D) segments are characterized by multi-modal mobility offers, short distances and low speeds (Schmitt and Gollnick, 2016). Short-distance transfers within airports provide the interface between feeder traffic and air transport (Schmitt and Gollnick, 2016). These are kerb-to-gate (K2G) and gate-to-kerb (G2K) segments. The flight component is operated by airlines in the gate-to-gate (G2G) segment. Passengers on connecting flights may pass through these segments several times. Business-to-business (B2B) suppliers, such as automotive and aircraft manufacturers, support TSPs. On the horizontal level, digital travel platforms like Google, Booking.com and Skyscanner serve as D2D mobility integrators in all five segments (Javornik et al., 2018; Schulz et al., 2018).

D2D travel is often examined in urban mobility studies, but these often neglect D2D air-travel chains. The latter include the air transport leg, which gives rise to increased research complexity for several reasons, including the difficulties of collecting real D2D air-travel-time data from passengers (García-Albertos et al., 2017) and combining travel legs in research activities such as surveys (Susilo et al., 2017). This study adopts a D2D perspective on air travel, focusing on the European market. The analysis covers TSPs operating in each of the five main segments, which together comprise the scope of D2D air travel.

2.2. Hypothesis development

In this section, seven key hypotheses are proposed, which are as mutually exclusive and collectively exhaustive as possible (MECE principle). These are based on previous literature, including (1) the results of a Delphi study of the future of D2D air travel (Kluge et al., 2020); (2) current demand-side pain points along the D2D travel chain (Baumgartner et al., 2016; Budd et al., 2016; Kim and Park, 2016; Monmousseau et al., 2019; Rothfeld et al., 2019; Sezgen et al., 2019; Ureta et al., 2017); and (3) disruptive trends in the aviation and overall mobility sector, such as the climate debate and exogenous shocks. As in other studies, scientific publications are considered as sources identifying the latest developments and most pressing challenges (Li et al., 2019). The trends characterized by the hypotheses are already affecting all or parts of the D2D air-travel chain. Hence, they must be understood and integrated into mobility service providers decision-making processes in order to offer integrated and successful D2D solutions. In other words, these hypotheses address the question: *what factors are important for creating successful D2D products and services?*

1. Personalization

Different passenger segments have differing needs (Kluge et al., 2018b). Different generations, such as millennials or young travelers (Garikapati et al., 2016) and the elderly (Siren and Haustein, 2013; 2015), have increasingly differentiated and fragmented customer needs that mobility service providers along the travel chain must be able to fulfill. We argue that it is essential to understand passenger types and provide personalized options to cater to all passengers needs. The personalization of travel (and other industries) has already begun and is predicted to shape the future of D2D air travel (Kluge et al., 2020). This

[1] For similar definitions and models, see García-Albertos et al. (2017) and Ureta at el. (2017).

[2] The kerb refers to the pedestrian area outside the terminal building.

Table 1
Hypotheses and related keywords.

Seg.	Hypotheses	Metrics (Keywords)
D2D	Personalization (H1)	Customisation, customisation, flexibility, individualisation, Mobility as a Service (MaaS), on-demand, pattern recognition, passenger, personalisation, preference, passenger segment, passenger need, passenger requirement, targeted advertising
D2D	Passenger data (H2)	Automated self-boarding, blockchain, biometric identity, customer data, chatbot, data, data mining, digitalisation, digital element, digital experience, digital service, digital journey, home-printed tag, mixed reality, mobile service, passenger data, predictive analytics, passenger mobile app, passenger experience, robot, autonomous machine, SITA, self-service, tracking, virtual agent
D2D	Establishing partnerships (H3)	Add-on product, Amazon, Artificial Intelligence (AI), Airbnb, air-rail connection, booking.com, bundled services, C Teleport, Culture Trip, co-modal, collaboration, cooperation, Dohop, Dreamlines, door-to-door, (D2D), Evaneos, Flightsayer (Lumo), GIVT, GetYourGuide, Google, HomeToGo, integrated service, IATA, integrated solution, interface, intermodal, interoperability, LuckyTrip, Maas Global, multimodal, Mobian, Netflix, nodes, Omio, partnership, Resolver, Skyscanner, Selina, Secret Escapes, TravelPerk, Tiqets, wemovo, WeSki, (customized keywords[4])
K2G. G2G. G2K	Environmentally-friendly air transport (H4)	Alternative fuel, ACA, carbon footprint, carbon offsetting, (carbon) compensation, CORSIA, decarbonisation, eco-friendly, ETS, electrification, emission, emission trading system, environment-friendly, environmental awareness, flight shame, Fridays for Future, green mobility, Gevo, Greenpeace, hydrogen, Neste, renewable fuel, renewable diesel, renewable aviation fuel, renewable drop-in kerosene alternative, SG Preston, Swedish Biofuels, sustainable energy system, sustainable aviation fuel
D2K. K2D	Airport feeders (H5)	Access, egress, accessibility, autonomous driving, BeeRides, Bipi, BlaBlaCar, Bolt, Blue Label, Car Sharing, car&away, Cabify, Cluno, feed, Free Now, FlixBus, innovation, last mile, Lyft, mobility concept, moovel, MyTaxi, novel / emerging mobility concept, on-demand private shuttle, pooling, ride hailing, RideLink, SeaBubbles, seamless, shared mobility, Sono Motors, Taxi2airport, transrapid, Uber, Viruo, Welcome Pickups
D2D	Disruption management (H6)	Automatic rebooking, alert, cascading effect, customer care, disturbance, disruption, disruption management, journey re-configuration, network congestion management, push notifications, real time status, service delay, travel assistant, travel flow, travel itinerary
D2D	Exogenous shocks (H7)	Crisis, exogenous shock, epidemic, hygiene measures, health on board, health screening, luggage screening, monitoring, passenger safety measurement, pandemic, resilience, reaction, recover, regulation, supply shock, security, technological shock, terrorist attack, terrorism, WHO

applies to ancillary products, such as pre-travel information, on-board services and in-flight entertainment, as well as to the core travel product itself. Based on these arguments, the first hypothesis is:

H1: Door-to-door air travel chains are highly personalized.

2. Passenger data

Providing enhanced, digital D2D travel services requires the use of passenger data. Airlines already possess large amounts of passenger-related information, such as demographic information, flight class (premium versus business), trip purpose and financial information (Javornik et al., 2018). It is predicted that such data will be leveraged to improve services along the air-travel value chain (Wade et al., 2020). Studies show that passengers are willing to share personal data if they receive benefits in return. For instance, 80% of time spent in the K2G segment is buffer time (Ureta et al., 2017), which might be reduced by improving airport processes and reducing waiting times (Monmousseau et al., 2019). Two-thirds of passengers would share more personal data to support faster airport processes (IATA, 2019). Digitizing the travel chain and leveraging passengers data may support the provision of integrated and digital D2D travel solutions, and help to predict future needs. This leads to the second hypothesis:

H2: Passengers data are leveraged to improve the digital travel experience and offer enhanced services along the D2D air-travel chain.

3. Establishing partnerships

Mobility service providers operate within their own travel segments and cannot control the services of other segments. To provide integrated and bundled services, partnerships and collaborations with other providers along the D2D air-travel value chain are essential. Examples include Rail&Fly (Lufthansa, 2019) and D2D luggage delivery services (Luggage Free, 2020). We argue that such collaborations may create a competitive advantage, as already proven in early studies of integrated air-bus products (Merkert et al., 2020). TSPs can become real D2D mobility providers. Collaborations with tech companies may enhance offers along the travel chain (Kluge et al., 2018a), as they have the necessary infrastructure and can provide customized content on personal devices, travel recommendations and other services. Hence, the third hypothesis is:

H3: Establishing partnerships with other mobility service providers and tech companies is essential for offering bundled D2D solutions.

4. Environmentally-friendly air transport

In light of current environmental debates and climate-related challenges, mobility must urgently change, for instance by decarbonizing transport, introducing carbon-offsetting programs and using renewable aviation fuels (RAF). Regulations and political agendas such as Flightpath2050 (European Commission, 2011), the EU Emissions Trading System (European Commission, 2020) and the Carbon Offsetting and Reduction Scheme for International Aviation (ICAO, 2020) are pushing airlines to explore alternatives to kerosene, including renewable drop-in fuels, hydrogen and electricity (Bauen et al., 2020). Programs fostering carbon reduction also apply to airports, such as the Airport Carbon Accreditation (ACA) scheme (ACI EUROPE, 2009). Environmental awareness and related behavioral changes appear to be increasing among the European population (European Commission, 2019), for example with regard to their willingness to reduce air travel, pay for carbon-offsetting and use environmentally-friendly substitutes for flights (European Investment Fund, 2020). TSPs need to react to this emerging trend on the demand side and create sustainable, D2D transport solutions, not only to remain competitive, but also to contribute to preserving the planet. The fourth hypothesis is as follows:

H4: Offering environmentally-friendly air transport solutions and related transport products is indispensable.

5. Airport feeders

The time taken to access and egress airports (D2K and K2D segments) needs to be reduced significantly. Using Google Maps data for 22 major European airports, Rothfeld et al. (2019) show that driving a private vehicle is still faster than using public transport. Alternative transport modes that bypass traffic jams (like subways) might also save travel time (Monmousseau et al., 2019). In addition to conventional airport feeders (Budd et al., 2016), some studies explore new transport concepts as alternative modes of D2D connections in Europe, such as on-demand air taxis (Sun et al., 2018)[3]. In addition to potentially autonomous air taxi solutions, other mobility concepts have already entered the market, such

[3] A recent study (McKinsey & Company, 2021) predicts a general shift in the European market away from private vehicles towards new modes of transport until 2030. Business travelers and high-net-worth individuals are predicted to be first movers for air taxis (Kluge et al., 2019). Although explored in the extant literature, we do not consider on-demand air taxis as a form of mass transport for airport feeder traffic, and hence exclude these and similar companies from this analysis.

as ride hailing and car-sharing, which are further changing passengers travel needs and mobility patterns (Bardhi and Eckhardt, 2012; Young and Farber, 2019). Enhancing the quality of airport feeds with regard to time and available modes improves the entire D2D travel experience. Hence, the next hypothesis is:

H5: Novel mobility concepts improve airport feeder traffic.

6. Disruption management

Over 55% of air passengers faced travel disruption in 2019 (IATA, 2019). This may occur when using airport feeder services, so many passengers arrive at the airport early to allow a time buffer, which accounts for up to 80% of time spent at airports (Ureta et al., 2017). Flight delays are another potential source of disruption, and may lead to negative emotions for air travelers (Kim and Park, 2016). Another disrupter is luggage delays (Sezgen et al., 2019). Delayed airport feeds, late flights and delayed luggage may impact negatively on onward connections and bookings in later travel segments. Disruption management entails avoiding disruptions to ones own services and, if necessary, acting upon them, for example through automatic rebookings, real-time flight status, alerts and luggage information (IATA, 2019). This may significantly improve passengers experience, saving them time and reducing stress. Hence, the sixth hypothesis is:

H6: Managing disruptions (delays and operations) along the D2D air-travel chain improves the passenger experience.

7. Exogenous shocks

Exogenous shocks are defined here as events that seriously affect the air transport system, including D2D air travel. Terrorist attacks and pandemics are examples. In recent years, terrorist attacks have occurred around the world (Esri and Peacetechlab, 2020), impacting on D2D air travel. For example, following the attacks of September 11, 2001 in the US, global air transport operations changed fundamentally with regard to airport security procedures and luggage screening (Blalock et al., 2007; TSA, 30.12.2002). A business travel stress model shows that personal safety, before and during trips, may also be a stressor at the individual passenger level, leading to fear and anxiety (Ivancevich et al., 2003). The long-term consequences of the current COVID-19 crisis for the transport and aviation industry are already obvious (OAG, 2020; Pearce, 14.4.2020; Škare et al., 2020). Such incidents may impact not only on demand for air travel, but also on passenger requirements and travel procedures, such as airport health screening during the Ebola epidemic (Gold et al., 2019). Passengers concerns about picking up infections during travel were identified some time ago (Gustafson, 2014). TSPs may be required to introduce additional health screening (Gold et al., 2019) and hygiene measures in cabins (IATA, 2020a). Exogenous shocks require mobility service providers to be resilient (Linden, 2021) and hence constantly adapt to new health and safety regulations, leading to the final hypothesis:

H7: Exogenous shocks (e.g. terroristic attacks and pandemics) require mobility service providers to constantly adapt towards new health and security issues.

Table 1 summarizes all the hypotheses and the keywords used to describe them. Both the hypotheses and the keywords were pre-tested with two aviation experts. The most frequently-used keywords were included, avoiding the provision of more detail than necessary for the models. The keywords contain essential information, and examples were retrieved from the trend hypotheses and company names, such as potential partners, start-ups and associations (see Table 3 and Table 10 for details). In the analysis, airports, airlines, public transport and railway companies were also included as keywords. These are listed in the Appendix (see Table 7, Table 8 and Table 9). Given the definition of D2D air travel in Section 2.1, the travel segments to which the hypotheses are most applicable are also indicated (see Seg.).

Table 2
Text analytics in transportation research.

Author(s), Year	Study Scope	Used Data	Methodology
Sezgen, Mason & Mayer (2019)	Investigation of key drivers for airline passenger satisfaction	OCRs	Latent Semantic allocation (LSA) (text mining and categorization)
Korfiatis et al. (2019)	Making airline service quality measurable in its various dimensions	OCRs	Structural topic models (STM) (extension to LDA method[5])
Lucini et al. (2020)	Measuring airline customer satisfaction	OCRs	Latent Dirichlet allocation (LDA)
Punel & Ermajung (2018)	Deriving passenger segments	Twitter data	Cluster analysis
Kim et al. (2017)	Measuring tourists' perception on travel experience	Online traveler reviews	Sentiment analysis
Kuhn (2018)	Detection of known and unknown trends and topics	Aviation safety reports	STM
Das et al. (2020)	Deriving main underlying research topics	Transport-related journal papers	LDA & STM
Lin et al. (2018)	Identifying key components of airline mission statements	Mission statements	Content analysis
Law & Breznik (2018)	Detection of embedded key values	Mission statements	Content analysis, network analysis
Seo (2020)	Detection of shared values within airline alliances	Mission statements	Content analysis
Kumar & Rao (2019)	Development of a performance measuring metric for airlines	Annual and business reports	Various statistical analyses
McLachlan et al. (2018)	Assessing the environmental impact on customers' decision making	Annual, sustainability and CSR reports	Consumer questionnaire

Table 3
Databases used in the analysis.

Seg.	Source Name	Data Extracted	Year
D2K, K2D	EUROSTAT	Top feeder traffic providers (public transport providers): extracted manually for top 20 European urban areas by inhabitants, including cities and commuting zones	2017
K2G, G2K	OAG	Top airports: largest European airports (by total planned seat capacity, for both arrivals and departures)	2018
G2G	OAG	Top airlines: largest European airlines (by total planned seat capacity)	2018
G2G	Mordor Intelligence	Top five players in renewable aviation fuels market (used as keywords)	2020
D2D	Crunchbase	Top transport start-ups by estimated revenue range (used as keywords)	2020
D2D	Lufthansa Innovation Hub	Europes 20 top-funded travel and mobility tech startups and corporate venture capital investments by European carriers, by total number of investments (used as keywords)	2020

2.3. Using textual data for transportation research

Analysis of textual data has several advantages, including cost efficiency, widespread availability of publicly accessible data from the Internet (Rose and Lennerholt, 2017), and ability to analyze large volumes of documents (Kadhim, 2019). Text analytics, as part of machine learning, is an interdisciplinary research method applicable to textual

[4] Customized keywords depending on TSP: airline, airport, railway, bus, public transport providers (see Appendix).
[5] Also allows to run correlations among developed topics.

data, grounded in the theory of content analysis. It is applied to unstructured data such as blogs and technical documents (Anandarajan et al., 2019), and semi-structured data such as social media content (Kinra et al., 2019). The Internet is the largest repository of research data (Rose and Lennerholt, 2017), although corporate data is increasingly being used in the industry (Bloomberg Professional Services, 2018; IDG Business Media, 2020).

Transportation research is increasingly using such data resources, owing to recent methodological advances and the variety of textual resources available (Kinra et al., 2019). This may reveal various insights, depending on the technique applied and the data resources available. This section provides an overview of how textual data can be used in transportation research, with a focus on aviation, as summarized in Table 2.

Researchers are applying text analytics techniques to user-generated online content, such as online customer reviews (OCRs). Sezgen et al. (2019) explore the drivers of passengers satisfaction with airline products and services using latent Semantic analysis (LSA) of passengers reviews retrieved from the travel platform TripAdvisor. Other studies follow a similar approach, using OCRs as data samples for text analytics (Korfiatis et al., 2019; Lucini et al., 2020)[6]. Textual data can also be used for market segmentation purposes. For example, Punel and Ermagun (2018) apply cluster analysis to Twitter data to explore airline passenger segments. Kim et al. (2017) provide a comprehensive overview of OCR studies in the broader context of tourism and hospitality, and apply sentiment analysis to online reviews by travelers visiting Paris. Kuhn (2018) applies the structural topic model (STM) technique to aviation safety reports to identify trends and topics in that field, and Das et al. (2020) use STM and latent Dirichlet allocation (LDA) to detect 20 key research topics in data drawn from transport-related research published in the *Transportation Research Record: Journal of the Transportation Research Board*.

Organizational textual data have also proven to be suited to text analytics. Law and Breznik (2018) apply content analysis and network analysis to airline mission statements to identify airlines embedded values, and Lin et al. (2018) use airlines mission statements to develop a framework of mission statement components. Seo (2020) detects shared values between airline alliance members by analyzing and comparing the content of 61 airlines mission statements. Annual reports have also been used as a data source. Kumar and Rao (2019) develop a performance evaluation instrument using information from airlines annual reports and financial statements. Secondary information on airlines can be collected by combining annual reports with sustainability and corporate social responsibility (CSR) reports, as in McLachlan et al. (2018) who study the impact of environmental concerns on passengers airline selection.

Textual analyses of OCRs have become dominant in research and can be used as a resource for organizational decision making, supplementing traditional customer surveys. OCR studies can use large sample sizes and can be updated daily, although their application presents some challenges (Roberts et al., 2014; Silge and Robinson, 2017). For example, funding is needed to retrieve a critical amount of historical data for research purposes (Kinra et al., 2019). Internal corporate textual data, such as emails, intranets and wikis, are insightful but not publicly accessible (Kinra et al., 2019), whereas corporate mission statements are publicly disclosed, conveying corporate values and visions. However, mission statements were considered too high-level to achieve the research objective of this study, which used annual reports and, if available and not already included, sustainability reports as data sources.

[6] The review of OCRs studies is not exhausted, this paper discusses most recent work.

2.4. Contribution to the literature and practical implications

Previous work has explored research questions using text analytics techniques, and has focused on parts of the D2D journey. Given the high practical relevance of intermodal, D2D air-travel solutions, this study focuses on the entire D2D air-travel value chain. It makes several contributions to the literature and has practical implications for the industry. First, it investigates a novel research context in its entire D2D scope, including all TSPs along the travel chain. Second, it conducts trend analysis, adopting an innovative approach to assess systematically whether companies consider and communicate key hypotheses and trends, using real-world data. Third, a trend dictionary is built and a text classification model developed, and this learning can be applied to other reports and Big Data, increasing productivity and saving time and other costs. Finally, the study applies text classification models in a novel research setting, and expands the field of application of text analytics and machine learning models.

2.5. Expected results

This review of D2D air travel indicates that most topics apply more to airlines and airports than to feeder traffic providers. D2D travel is relevant to all customers of airlines and airports, but to only a fraction of feeder traffic providers customers. Feeder traffic providers also provide public transport for urban passengers, which accounts for a large part of their operations. Therefore, we expect most hypotheses to be more relevant to airports and airlines, as all of their customers need to access the airport and leave for their final destinations. Three percent of the EUs greenhouse gas emissions are generated by the aviation sector (European Commission, 2020), which also generates other emissions (Lee et al., 2020). This might be reduced by, for instance, switching to RAFs, although airports must also provide the required infrastructure. As airlines and airports are making the biggest efforts to achieve environmentally-friendly transport, H4 (Environmentally-friendly air transport) focuses specifically on the K2G, G2G and G2K segments. Hence, we expect this topic to be of greater relevance in airlines and airports communications, with an even stronger focus in airlines textual data. Improvements to airport feeder traffic (H5) affect mainly the D2K and K2D segments; hence, we expect this hypothesis to be particularly relevant in documents from feeder traffic providers. H6 (Disruption management) and H7 (Exogenous shocks) apply to all three providers equally. Overall, we expect higher results for airlines and airports, and lower results for the feeder sub-sample (see Fig. 2).

3. Methodology

Using text classification models, this study explores whether TSPs along the D2D air-travel chain consider key trends and findings from research to be strategically relevant, and hence include them in their communications. This section elaborates on the methodological approach and data preparation.

3.1. Research objectives and approach

For the purpose of this study, text classification models were built to analyze unstructured textual data to examine the relevance of the identified trends. In these models, pre-defined hypotheses, described by keywords and later by training texts, were assigned to corporate textual data, allowing documents to be allocated to predefined classes (supervised learning) (Grimmer and Brandon, 2013; Kadhim, 2019). These a priori defined classes were developed in the form of hypotheses (see Section 2.2) following a deductive approach (Welbers et al., 2017). The study followed a five-step approach (see Fig. 3), adapted from Rose and Lennerholt (2017) and Mirończuk and Protasiewicz (2018).

Table 4
Overview of data sources.

TSPs	Companies in Sample	No. of Annual Reports	No. of Sustainability Reports
Feeders	26	15	8
Airports	22	18	9
Airlines	23	14	10
Sum	71	47	27
Object of analysis	74 reports merged to **52 documents** (representing 71 TSPs)		

3.2. Data acquisition

Corporate textual data were used as data sources, the successful combination of which has been proven in previous studies (McLachlan et al., 2018). Annual reports are publicly disclosed, published yearly and retrievable from corporate websites. In addition to financial disclosures, they contain non-financial information (Esendemirli, 2014) on strategy, marketing, and research and development, making this published content strategically important. If not already contained within annual reports, any available sustainability reports, and documents with similar sustainability-related content such as CSR reports, were included for analysis. To create the desired D2D air travel scope, data from airports, airlines and feeder traffic providers were included in the sample. The largest European airlines and airports were taken from the Official Airline Guide (OAG, 2018). Airlines that were part of an alliance were considered individually. Airport feeder traffic providers operating in the most highly populated urban areas (by total number of inhabitants) in Europe (Eurostat, 19.3.2020) were researched manually. Table 3 lists the databases used to develop the rankings, and other sources used to detect keywords (Crunchbase, 2020; Lufthansa Innovation Hub, 2020a; 2020b; Mordor Intelligence, 2020).

Using these rankings, a data overview table was created, and the top airlines, airports and feeder traffic providers were researched. The final sample contained data from 71 mobility service providers along the D2D air-travel value chain, comprising 23 airlines, 22 airports and 26 feeder traffic providers (for a full list, see supplementary material). Several transport providers belonged to a single group, and are hence presented in group reports, such as the International Airlines Group (IAG), Aroports de Paris (ADP) Group and Deutsche Bahn (DB). As the entire group reports were analyzed, the sample contained additional transport providers that were part of these groups but not part of the sample. Hence, the sample size was actually larger. The 71 TSPs in the sample were represented by 47 individual annual reports (or documents with similar content, such as company presentations and activity reports) and 27 sustainability reports (or documents with similar content) (see Table 4). These were downloaded manually from corporate websites in PDF format in English in April 2020. Detailed lists of the sample are presented in the Appendix. The annual reports were merged with any sustainability-related reports for the same company or group. Acquiring annual reports from mobility providers corporate websites has proven successful in previous research (Kumar and Rao, 2019). The base year was mostly 2018 or 2019, depending on availability[7]. Sustainability-related reports are not published yearly and might therefore be older.

The airline sample represented 69% of the total seats offered to and from Europe by European airlines, including both low-cost and full-service network carriers. The airport sample represented 40% of total seats for departures and arrivals at European airports (OAG, 2018). If reports were unavailable (e.g. non-functioning URL), were not translated into English, or were read-only, the company was excluded from the sample and the next entity on the list was used until a sample size of at least 40% market coverage in Europe was reached. Such sample sizes are considered to provide insightful results. The dataset can be considered to be “Small Data” (versus Big Data). Small Data may also produce valid and insightful findings, while reducing costs and resources (Faraway and Augustin, 2018). This study aimed to conduct analysis at an aggregated TSP level. As a sufficient sample had already been gathered at a market coverage of 40%, additional data would have not improved the findings, but would have incurred additional expense.

[7] For some companies, the financial years ends in March.

3.3. Data preparation

The raw data were uploaded into R and saved in corpus format, a collection of associated texts that is easy to manage (Benoit et al., 2018). Separate corpora were built, representing airport data (N=18), airline data (N=18), feeder data (N=16) and the entire sample (N=52). Raw data must be reduced in volume and complexity, manipulated and pre-processed to make them usable for learning methods. To keep the data volume low, all PDFs were converted into plain.txt format and encoded into UTF-8 standard, as recommended by Grimmer and Brandon (2013) and Welbers et al. (2017) and as required for application of the *quanteda* R-package (Benoit, 2018).

Documents contain noise, such as unnecessary stop words and special characters (Mirończuk and Protasiewicz, 2018; Welbers et al., 2017). These were removed during the data preparation, following Welbers et al. (2017) five-step guideline. This process comprised: (1) importation of the textual data into R as raw text corpora, allowing analysis of the data at a document level; (2) string operations; (3) pre-processing of the data, including tokenization, normalization and stemming, and removing stop words using the default list in the *stopwords* package in R (Benoit et al., 2020) and the authors customized list[8]; (4) creating a document-term matrix (DTM) using the *quanteda* package (version 2.1.2), which also included some mentioned data preparation steps (Benoit et al., 2018); and (5) filtering and weighting the DTM. Descriptive statistics were produced to explore the data, such as filtering out the top features in the sample and sub-samples, and creating word clouds (see Fig. 4).

The raw text corpora differed considerably (see Fig. 5). The airport data comprised a much higher number of tokens, although conclusions cannot be drawn about the content. After pre-processing the data, this imbalance disappeared[9].

4. Development of classification models

The classification models for this study were developed in a multi-labeled setting (as opposed to multi-class classification (Sokolova and Lapalme, 2009)). Hence, each document might belong to more than one pre-defined class. As described above, the seven classes for the text classification problem were: 1) Personalization, 2) Passenger data, 3) Establishing partnerships, 4) Environmentally-friendly air transport, 5) Airport feeders, 6) Disruption management, and 7) Exogenous shocks. Several approaches are available for classifying documents into multiple categories that are known a priori. Two were applied and compared in this study: dictionaries, and supervised learning classification models (Grimmer and Brandon, 2013).

4.1. Dictionary-based classification model

4.1.1. Building the dictionary

Dictionary analysis counts the frequency of keywords in a sample (Watanabe and Zhou, 2020). In this study, a dictionary was developed

[8] *Lemmatization* was also considered to improve the results, but according to Welbers et al. (2017), stemming is often sufficient in the English language. Words with similar meanings were included.

[9] The total numbers of tokens were 615,205 for the airport sample, 633,005 for the airline sample and 461,388 for the feeder sample.

based on the seven hypotheses (hence seven classes) and the knowledge-based keywords identified in the literature review (see Section 2.2), using the *quanteda* dictionary function (Benoit, 2018). Each term was used only once for each hypothesis. A basic dictionary of 225 entries was built. Each TSP was also tested for 127 customized keywords[10], leading to an average dictionary of 352 keywords. The keywords were stemmed using the value type *glob* style.

The data sample varied in terms of token volume per document and number of documents per sub-sample. To correct this imbalance, the dictionary results were weighted to calculate the relative frequency (see Eqn. 1). For each document, the number of hits per class was divided by the total number of tokens in each document. The sum of all documents was then divided by the total number of documents (*N*) in the respective sub-sample (airlines, airports or feeder traffic providers). Relative dictionary weightings have also been applied in previous studies as described by Watanabe and Zhou (2020).

$$\text{Relative frequency (per class)} = 100 \times \left(\frac{\sum_{i=1}^{N} \frac{\text{hits in document}}{\text{tokens in document}}}{\text{N}} \right) [\%] \quad (1)$$

where:

i = document's number in subsample

N = amount of all documents in subsample

4.1.2. Interim results

After applying the dictionary, the interim results (see Fig. 6) gave an early indication of which hypotheses (keywords) might occur in the TSPs documents, and which might be less represented. The dictionary matched around 7% of the textual data in the sample. The results indicated that H1 (Personalization) and H4 (Environmentally-friendly air transport) exhibited the highest matches. H2 (Passenger data), H5 (Airport feeders), H6 (Disruption management) and H7 (Exogenous shocks) were each represented with similar frequency counts in the data, suggesting that these four themes might be similarly relevant to TSPs communications efforts. H3 (Establishing partnerships) was least represented. However, a strong effect appeared after the customized keywords were applied at the TSP level in each dedicated sub-dictionary. Without testing for other TSPs names, H3 matched 4499 data points (versus 14,173). Hence, the dictionary classifier was very sensitive to providers names. The results also provided frequency counts for each TSP level.

Terms that seemed to be too general in meaning (e.g. *journey, travel*, mobi*, produc*, system*, aviat*, effect*, mana*, technol*) were excluded before applying the dictionary. Exploration of individual entries in the interim results revealed that some terms matched with many data points, such as *passeng** (passenger) (H1), *servic** (service) (H1), *board* (board) (H2), *integr** (integration) (H3), *sustain** (sustainability) (H4), *energi** (energy) (H4), *share** (sharing) (H5), and *regul** (regulations) (H7). All of these also seemed quite general, making it difficult to generate a high discrimination value for the respective hypothesis.

4.1.3. Validation

One of Grimmer and Brandons (2013) four main principles of quantitative text analyses is the need to validate the results. For instance, errors may occur when applying dictionaries that are too generic or keywords developed in other contexts or outside the domain in focus. Validation was carried out by pre-testing the keywords and hypotheses outlined in Section 2.2 on two mobility researchers who were not involved in the study. As the dictionary was also based on scientific publications and databases (Table 3), a robust foundation was built using sources from a comparable domain to the data sample. Hence, errors due to applying a generic dictionary were avoided. The interim results also showed that around 7% of terms occurred in the data, indicating the relevance of the selected keywords. However, some terms showed high ambiguity as their meanings were general. These generated many more matches than the others. Furthermore, the dictionary method does not allow the retrieval of additional keywords to improve the analysis.

Overall, the dictionary answered the research question simply but efficiently. The task of counting the frequency of keywords was conducted using the dictionary function, which would have been costly to conduct by manually coding all reports by hand while ensuring intercoder reliability. However, this approach has some limitations, as described above. The next section explains the complementary analysis conducted to explore the data further, beyond the purely knowledge-based keywords from the dictionary classifier. The supervised learning approach enabled us to make class predictions, rather than looking only at frequency counts.

4.2. Supervised learning

4.2.1. Approach

A supervised learning model was developed for multi-labeled classification (Fig. 7). Three basic steps adapted from Grimmer and Brandon (2013) were followed: (1) construct the training set, select features and train the model; (2) apply supervised machine learning to the test (validation) dataset; and (3) validate the models output and classify the remaining dataset.

4.2.2. Construction of a training dataset and selection of features

Diverse textual sources were used to train the model and transform the training texts into additional features as inputs for the predictions. As the initial data sample was small, new hand-labeled texts outside the sample were used to construct the training set for each class, such as the keywords discussed in Section 2.2 and other related texts. In order to overcome the limitations of the dictionary method, domain-specific (real-world) textual data from corporate sources were leveraged, such as Emirates (2019), Geneva Airport (2018) and ÖBB (2019)[11]. Not all classes seemed to be well represented in real-world corporate texts, which was another reason for conducting this study. Hence, reports and scientific papers were added, including Li et al. (2019) and Wenzel et al. (2020), as well as reports such as those published by SITA (2019) and accenture (2018).

To avoid data bias, the training texts were evenly balanced between the seven classes. Each class was represented by between 1100 and 1500 unique tokens (raw data after pre-processing). Including additional training text and extracting features helped to detect more relevant keywords and exclude human bias from the knowledge-based keyword selection. The training dataset was pre-processed using 1-*gram*. To reduce noise and irrelevant words and to avoid overfitting, the training texts were trimmed in the feature selection phase using different term frequencies (minimum term frequencies of 10, 15 and 20). As shown in Fig. 8, the number of features decreased with higher term frequencies. The training texts were pruned to achieve a minimum term frequency of 20. A higher number would have resulted in fewer keywords than in the dictionary classifier, and would have been counterproductive for leveraging additional features. Moreover, the curve stagnated after the minimum term frequency of 20. The results were compared in the validation phase.

4.2.3. Application of classification models

The data sample was split into a test dataset (N=5; the first five

[10] The keyword list for H3 included additional airlines, airports and feeder traffic providers names, which were applied selectively, leading to three sub-dictionaries for airports, airlines and feeder traffic text corpora. The basic dictionary is shown in the Appendix.

[11] All not part of the sample.

Table 5
Difference scores for estimated class probability with NB versus SVM classifiers (min. term frequency for pruning training texts in feature selection phase in brackets).

	NB (10)	SVM (10)	NB (15)	SVM (15)	NB (20)	SVM (20)
Difference score	0.101	0.063	0.083	**0.057**	0.107	0.059

reports from the sample) and the remaining dataset (N=47). The *quanteda.textmodels* package (version 0.9.1) was used to build the classification model (Benoit et al., 2018). The multinomial naive Bayes[12] (NB) (Manning et al., 2008) and linear SVM[13] (Fan et al., 2008; Vapnik, 1995) classifiers were used to make predictions for the class probability.

Previous surveys have revealed the NB classifier to be a suitable and well-performing classifier for document classification (Ting et al., 2011). The linear kernel of the SVM is recommended for text classification (Kowalczyk, 2017). Each class was trained on textual data from the training set, representing content relating to the hypotheses in the seven predefined classes k. Both algorithms learnt from the training texts, as examples of input x for output y (Goodfellow et al., 2016). The results showed the estimated probability distributions for each document in each class, with documents able to belong to multiple classes (multi-labeled setting).

4.2.4. Validation of classification models

Various metrics are available for evaluating the performance of machine learning models (Anandarajan et al., 2019; Goodfellow et al., 2016; Sokolova and Lapalme, 2009). This model's multi-labeled setting and low predictions (meaning a class was not well predicted for a document) were essential to answer the research question. The model's outcomes should support decision making and save costs and resources, and there were no correct or incorrect predictions. No pre-labeled validated dataset was available, and hence no measurements and metrics (e.g. Anandarajan et al. (2019)) for a multi-labeled classification (Sokolova and Lapalme, 2009) were feasible for use in this case.

This section compares the results from the test and remaining datasets, and pre-processing of the training data using a difference score (see Eqn. 2). The applied score describes the difference between the average estimated class probabilities from the test and from the remaining dataset, as a measurement of convergence. A low number indicates better model performance as the test and remaining datasets provided similar results. Table 5 presents the results of the validation phase.

$$\text{Difference score} = \frac{\sum_{i=1}^{k} |\text{probability test} - \text{probability remain}|}{k} \quad (2)$$

where:

i = class number
k = amount of all predefined classes

Overall, the linear SVM classifier yielded the best model output with training texts pruned to a minimum term frequency of 15. These pruned training texts still incorporated at least 40% of the keywords from the basic dictionary (67% for H1, 43% for H2, 21% for H3, 43% for H4, 23% for H5, 43% for H6, and 48% for H7). The output is depicted in Fig. 9.

4.3. Dictionary-based classifier versus supervised learning

Both the dictionary-based classifier and the supervised learning model answered the research question and enabled the development of dynamic models, updatable with new keywords, training texts and classes. However, the results differed between the classifier outputs (see Fig. 6 and Fig. 9). Several hypotheses (e.g. H1 and H6) had high frequency count matches using the dictionary-based method but showed a low predicted class probability in the supervised learning model. Pruning the training texts may have excluded rare keywords initially developed at the beginning of the study. As explored in Section 4.1, some keywords had a high frequency count match but were quite general, making it difficult to generate a high discrimination value for the hypotheses. This may also have been a sign of low relevance, as they only had a low frequency in the training texts for the supervised learning model. This is a key limitation of the dictionary-based classifier.

Conversely, the advantage of the supervised learning method for the purpose of this study was the extraction of additional keywords to make predictions. This partly eliminated human bias from the keyword selection phase, although such bias might still have occurred in the selection of training texts. Including additional training texts might change the background and interpretation of the classes, as explored in Section 5. Table 6 presents a comparison of the two approaches. In summary, owing to its predominant advantages, the next section focuses on the supervised learning model using the SVM classifier.

Table 6
Dictionary-based classifier versus supervised learning model to answer research question.

Criteria	Dictionary-based Classifier	Supervised Learning Model
Answers research question	Yes	Yes
Keywords required	Yes	No
Textual training data required	No	Yes
Retrieves new keywords	No	Yes
Rare keywords excluded	No	Possible (pruned texts)
Humans bias	Yes	Partly
Changes class meaning	No	Partly
Dynamic model (updatable)	Yes	Yes

5. Results

This study aimed to answer the research question: *to what extent do TSPs consider prevailing factors in their strategic planning and incorporate them into their communications?* The results do not provide qualitative assessments, such as whether TSPs had positive or negative associations with the trends. Furthermore, the analysis was carried out on corporate documents containing high-level strategic management topics. Trends not communicated in these reports may already have been implemented operationally. Internal corporate textual data would be required to investigate further.

5.1. Trend clusters

The outcomes from the supervised learning classifier model using linear SVM are shown in Fig. 9. The results for all TSPs along the D2D air-travel chain show three main clusters (see Fig. 10):

1. **Trends with higher relevance:** environmentally-friendly air transport (H4)
2. **Trends with medium relevance:** disruption management (H6), passenger data (H2), and airport feeders (H5)
3. **Trends with lower relevance:** establishing partnerships (H3), personalization (H1), and exogenous shocks (H7).

Fig. 10 indicates no relationship between class probability and the number of terms per class in the training texts. The individual results for each TSP show large differences. The subsets of airport, airline and feeder data are explored in the next sections.

[12] prior = "termfreq", distribution = "Bernoulli".
[13] weight = "termfreq", using *LiblineaR* package built in *quanteda.textmodels* package.

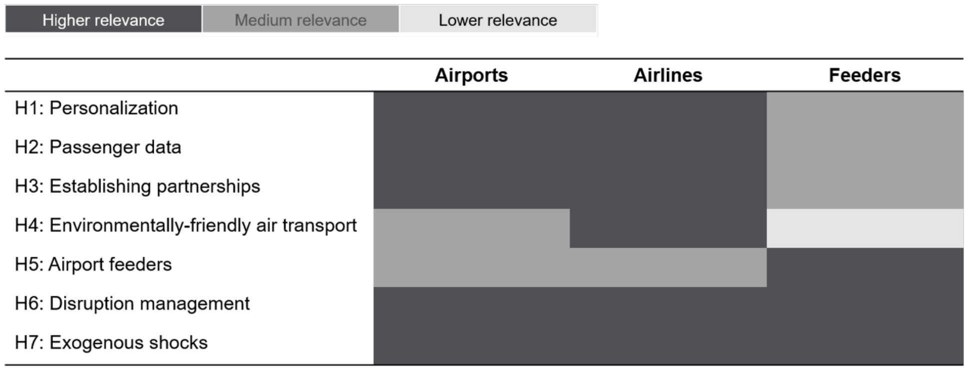

Fig. 2. Stakeholder analyses: Expected results at TSP level, from weak (light gray) to strong (dark gray) effects.

5.2. Airports

Fifteen reports remained in the airport subset after splitting the sample into test and remaining datasets. Measured by average values (means), H4 (Environmentally-friendly air transport) has by far the highest probability, at 0.57. One reason may be that airports are very active in operating in an environmentally-friendly way. As they are public service entities, they depend on public funding and tightly regulated. Thus, they have more need to publish on environmental issues. In the context of airports, and in light of the additional training texts, H4 may also encompass noise, animal protection, water management and recycling, as well as strategic partnerships with nature and animal conservation, which were not originally included in the hypothesis development and also relate to the content of H3 (Establishing partnerships). New keywords and topics may have arisen in the training texts. As seen in this example, classes may develop beyond their original context when training texts are included in the supervised learning model. In addition, the high relevance of H4 may also be grounded in the selection of the analyzed data, as 27 sustainability-related reports formed part of the overall sample.

H2 (Passenger data) and H6 (Disruption management) are also predicted with comparably high class probabilities, of 0.27 and 0.14 respectively, in the textual data for airports. Airports seem to focus on offering environmentally-friendly transport, leveraging passengers data and managing disruptions. Class probabilities of almost zero are predicted for the remaining classes. Hence, these key trends seem to be less frequently covered in airports textual data. The low class probability for H5 was expected (see Section 2).

5.3. Airlines

Two documents were used in the test sample, and the remaining 16 reports were included in the airline subsample. As shown in Fig. 9, a very

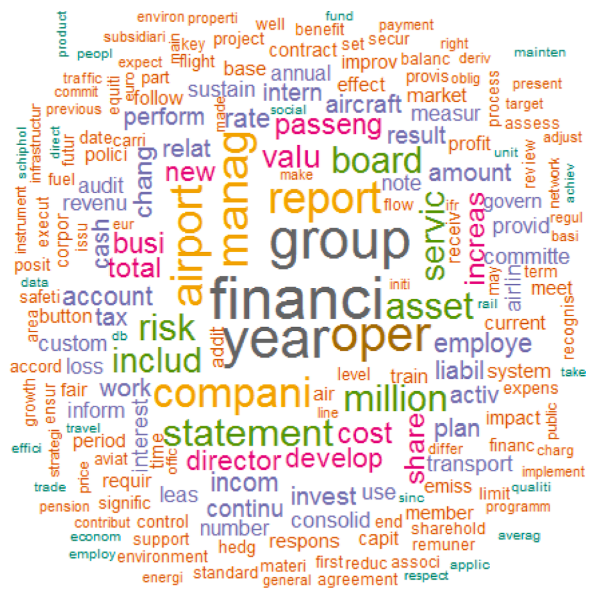

Fig. 4. Word cloud for top 200 words in entire sample (minimum term frequency 1,654).

high class probability of 0.96 for H4 (Environmentally-friendly air transport) is predicted by the SVM classifier model. This topics high relevance was expected for airlines textual data, but somewhat contradicts the industrys current damage to the environment through greenhouse gas emissions and other pollution. As this topic is of high strategic relevance and is driven by regulations and the public, it is unsurprising that airlines need to communicate about this challenge and explain themselves. Amongst various activities, airlines are investing in new

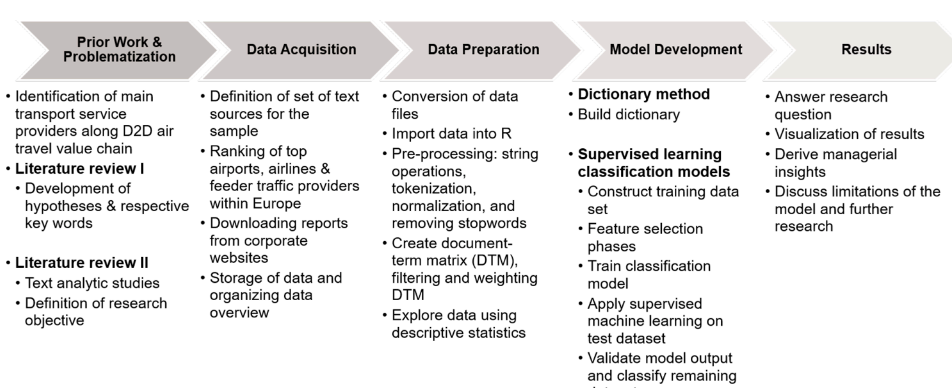

Fig. 3. Development of classification models: five-step approach (adapted from Rose and Lennerholt, 2017; Mironczuk and Protasiewicz, 2018).

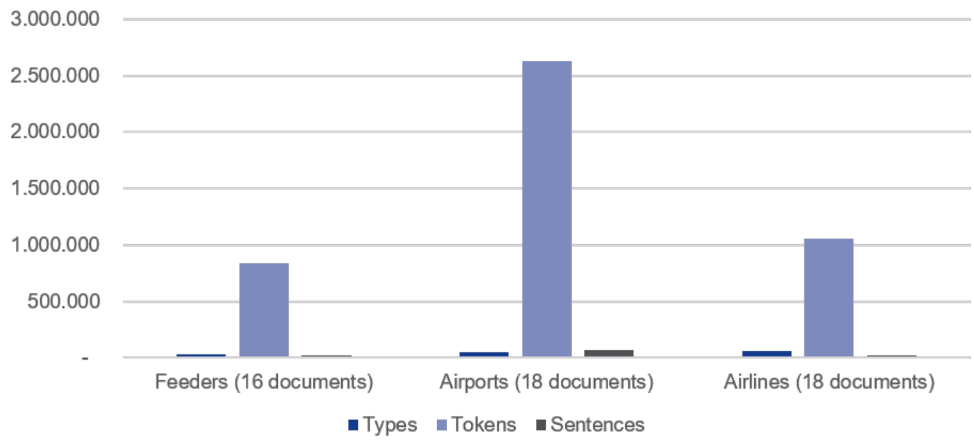

Fig. 5. Overview of raw data (before pre-processing).

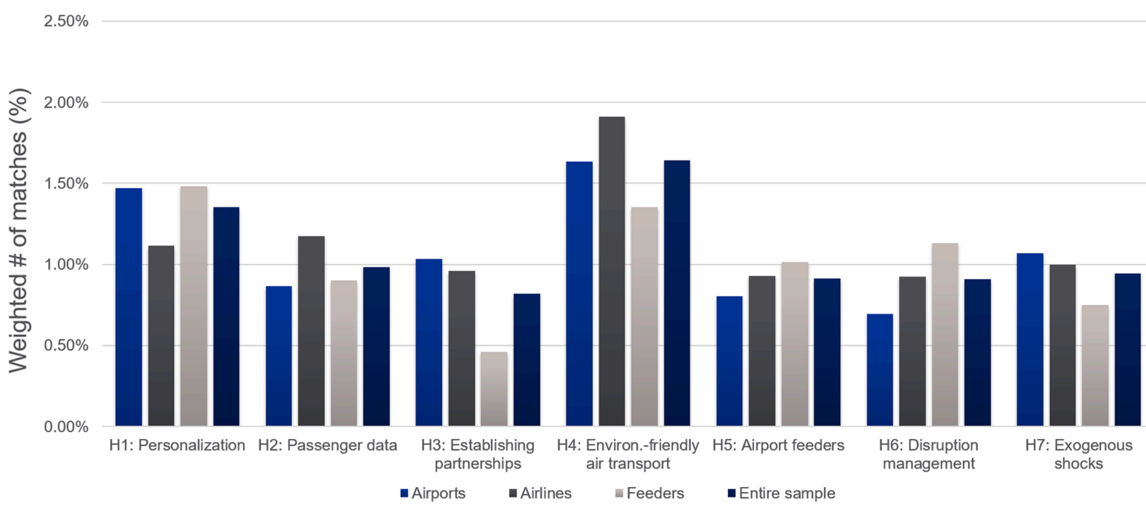

Fig. 6. Weighted results of applying dictionary method.

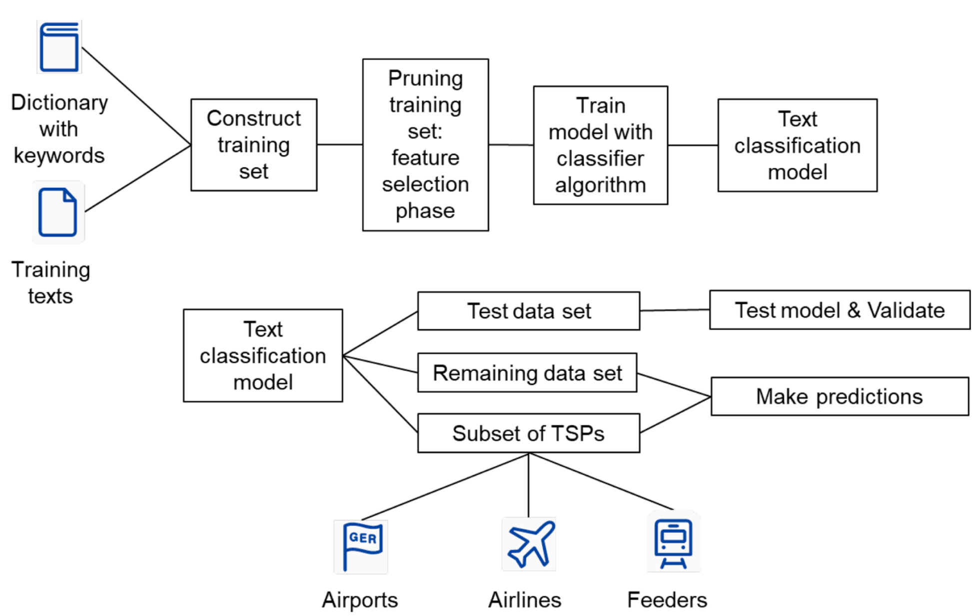

Fig. 7. Supervised learning model for text classification (authors depiction).

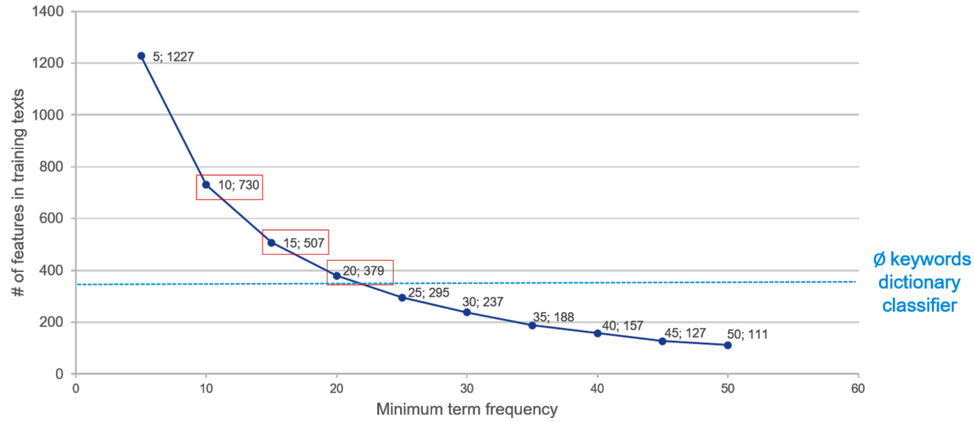

Fig. 8. Pruning the training texts versus number of features.

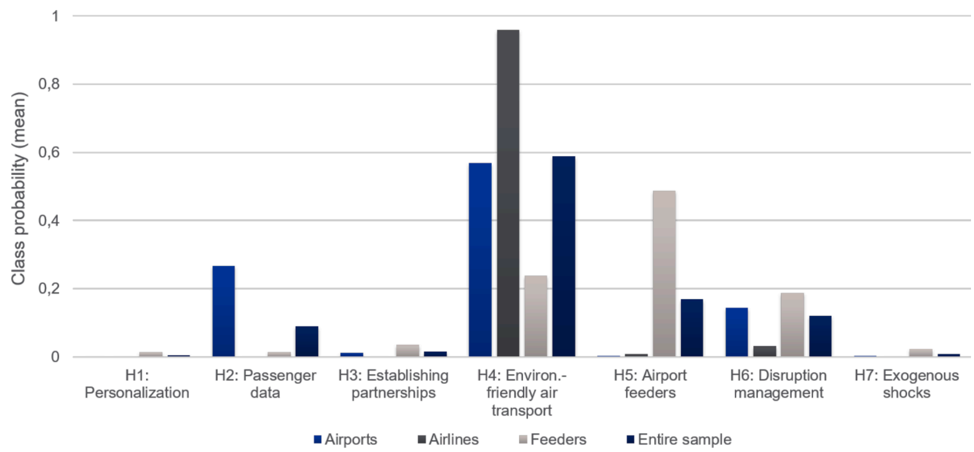

Fig. 9. Final results from classification model output applying linear SVM to remaining dataset (N= 47) at TSP level.

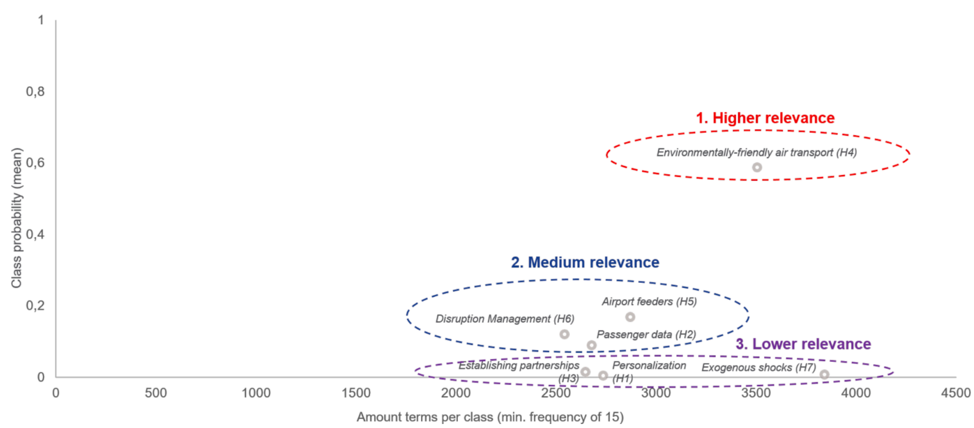

Fig. 10. Classification model output in three clusters.

fleets that reduce weight and fuel consumption. As previously discussed, this high prediction may also be grounded in a biased sample or the leveraging of additional training texts and the overall sample.

Low average class probabilities of 0.03 and 0.01 are predicted for H6 (Disruption management) and H5 (Airport feeders) respectively. The remaining hypotheses have a class probability of zero, and hence seem to be less frequently covered in airlines reports.

5.4. *Feeders*

The highest relevance in feeders data is indicated for H5 (Airport feeders), with an average class probability of 0.49. This class includes aspects like travel time, seamless and shared mobility, and on-demand solutions, as well as general mobility, such as urban, day-to-day transportation. It is hence unsurprising that feeder traffic providers show a high prediction for that class. Contrary to previous expectations, the second-highest class probability is predicted for H4, showing that environmentally-friendly transport may also be relevant and important for ground services covering the D2K and K2G segments. An average class probability of 0.19 is predicted for H6 (Disruption management), which is assumed to be critical for all three TSPs. Low class probabilities are predicted for H3 (Establishing partnerships) and H2 (Passenger data). Novel transport modes and mobility solutions also rely on leveraging passengers data and offering innovative mobility solutions based on predictive models. The sample included public transport providers, but excluded mobility provider companies such as Uber, Blacklane and Free Now. The results might differ with an extended feeder sample, so further research is required. As in the overall sample, lower class probabilities are predicted for H1 (Personalization) and H7 (Exogenous shocks). Personalized and tailored mobility solutions for urban air mobility (on-demand and time efficient) lie far in the future, as currently only prototypes are available. Hence, personalization of airport feeders may also relate more to the future, with less relevance in current corporate reports.

5.5. Managerial insights

Before elaborating on practical managerial insights, it is necessary to consider the overall context within which the results must be interpreted. Annual reports were used as the object of analysis in this study. These types of report are subject to disclosure requirements, depending on the stock exchange on which the company is listed. Traditionally, disclosures have been very clearly oriented towards regulations and the financial market, explaining how past financial results were achieved. However, the content of reports has evolved in recent years. For example, the contents of sustainability reports, which were also part of this analysis, relate increasingly to governance and environmental issues. It is therefore unsurprising that keywords in the area of strategy occurred less frequently in the data sample, whereas keywords driven by regulation occurred more frequently. For instance, the results reveal a high class prediction for H4, which is concerned with environmentally-friendly air transport. In summary, we expected topics relevant to top management to be included in the reports. Other aspects might already be established in companies, in the context of the operational business rather than as top management topics.

This study is of benefit to the industry in several ways. First, the hypotheses developed in Section 2.2 provide a comprehensive overview of current trends and drivers of D2D air travel. Given the high practical relevance of this topic, these insights will support strategy and decision making in organizations seeking to become real D2D mobility suppliers. For instance, partnerships (H3) are identified as a main driver of successful D2D products and services. In light of the current COVID-19 pandemic, potential partners might include suppliers of masks and hygiene products. Partnerships are also essential to mobility providers seeking to focus on seamless D2D air travel. Given that the main driver is personalization (H1), these might be unconventional partners, as personalization is already strong in other domains (such as tech companies). Amazon, Google, Netflix, Apple, Facebook and Spotify already possess customers data (H2) and could improve passengers personalized experiences throughout the travel chain, with regard to both actual travel and by-products. Of course, using personal data is only possible with passengers consent. However, research shows a general willingness to do so if there is a benefit in return.

From the classification results, it can be concluded that not all TSPs are communicating much with regard to personalization (H1) in the D2D air-travel chain, and hence may not consider this trend to be strategically relevant. With more diverse passenger profiles and needs (Kluge et al., 2018b), this will become increasingly important. Other travel providers (platforms) outside the aviation industry are already starting to provide customized options for traveling through access to passengers data (H2), potentially generating additional revenue. On the horizontal level, digital travel platforms like Google, Booking.com and Skyscanner may serve as D2D mobility integrators in all five segments. Platforms like EcoPassenger compare carbon dioxide emissions, energy resource consumption and other components of the entire travel chain to offer environment-conscious travelers the most environmentally-friendly options (EcoPassenger, 2020).

Alongside the low relevance of personalization (H1), airports show a medium level of relevance of passengers data (H2). As a Delphi-based scenario study recently revealed (Kluge et al., 2020), the future of D2D travel in the European market may be strongly influenced by these two key trends. To remain competitive, given the currently low demand for travel in many parts of the world, airlines are strongly advised to think about using passengers data to personalize their travel. Improving airport feeders (H5) might also allow the creation of personalized D2D travel chains, especially in light of the COVID-19 crisis. Passenger segments such as business travelers and elderly passengers may increasingly demand contact- and touchless travel chains to minimize face-to-face contact and infection risks.

The results also depict a low relevance of TSPs adaptation to exogenous shocks (H7). As recently observable in the market, aviation and mobility players need to learn to react much more flexibly to crises. TSPs must use this learned resilience in the next few years to meet passengers day-to-day needs. For instance, flexible booking and ticketing systems might be better adapted to current demand. This will become increasingly important in strategic decision making, as it is likely to take longer than expected for the travel industry to recover from the COVID-19 crisis (Škare et al., 2020). Improving airports access and egress (H5) is shown to be highly relevant to feeder traffic providers. Airports are advised to increase their partnerships (H3) and improve airport feeder traffic (H5) by leveraging novel mobility concepts. These two key trends are interrelated, as increasing partnerships with novel mobility companies may lead to improved feeder traffic for passengers. Conversely, established feeder providers in this sample may be suitable partners.

Finally, the text classification model approach may allow companies to be assessed more cheaply and quickly. The classification model can be used to determine whether a company is communicating on the most pressing trends. It may also support companies internal assessments of their own communications efforts, and help them to analyze stakeholders and competitors in the market. Other potential applications include mergers and acquisitions, the stock market, the investment sector, especially with respect to the often time-critical due diligence, and other consulting projects. In such deployments, one might start with a real-world business problem, investment level or stakeholder identification, and then determine data availability (Rochwerger, 2019).

6. Limitations, future research and conclusions

6.1. Limitations and future research

The limitations of this study offer ample opportunities for further research to advance both the prototype models and data collection. The dictionary method has several limitations, such as simple representations derived by counting the frequency of pre-defined keywords (Watanabe and Zhou, 2020), and the development of keywords outside the domain or with human bias (Grimmer and Brandon, 2013). In this study, keywords in the dictionary were knowledge-based and developed by mobility experts, providing some external validity. Although this approach fitted the overall research goal, it can also be seen as a limitation, as humans may still miss relevant keywords or use the wrong terms. Watanabe and Zhou (2020) describe unsupervised topic models to detect keywords from textual data, such as Bayesian hierarchical topic modeling and LDA. These alternative approaches might counteract the limitations.

To exclude human bias in this study, a training set was constructed in the second step of the classification model development, using mainly domain-specific textual data. As described by Grimmer and Brandon (2013), human coding should best develop iteratively, meaning that several people are involved in the coding exercise, in order to avoid ambiguities or missing texts. In this regard, the model could be improved, as in this study only one coder was used. When looking at the training texts in detail after the pre-processing, we discovered some garbage words with no meaning, such as *one, also* and *can*. These could be excluded beforehand by using a customized list. However, there were very few garbage words compared with the number of valuable words, and their effect on this automated approach is considered to have been small. Furthermore, additional classifiers might be applied in the supervised learning model, or pre-processing of the training dataset might be adapted, such as using 2-*gram* or 3-*gram* (Kowsari et al., 2019). To overcome annotation and data biases, additional or different training texts might be used, such as excluding all academic data and fitting the model with corporate textual data only.

This study examined the extent to which hypotheses are considered by companies, as revealed in their communications in corporate reports. Hence, this study can provide answers if a trend is relevant, but cannot provide qualitative assessments, for instance of whether a company has positive or negative associations with a trend, or whether it occurs in the context of that hypothesis. In future research, sentiment analyses might deliver more detailed insights. Image recognition might be another complementary next step, as company reports also communicate through visual images, such as trees, animals, new projects or products and passenger groups, that are not captured in the classification model.

The overall sample also had limitations. Not all organizations, such as start-ups or private companies, publish annual reports. Hence, start-ups had to be excluded as possible feeder traffic providers to be analyzed in this study, but were included as keywords. As already touched on, data biases may also have been present in the data sample itself. We included many sustainability reports in the sample, so strong results for H4 may be unsurprising. Using company reports as data sources has limitations owing to content variance. There is no consensus on understandings of corporate sustainability in corporate social reports (Landrum and Ohsowski, 2018), and some companies disclose more information than others, making it difficult to draw comparisons at a company level. There were also large variances in the raw data for different types of TSP. For instance, although the numbers of documents for all three subgroups were similar, the number of tokens within these documents varied greatly. After pre-processing the data, this imbalance disappeared, showing the critical importance of this research step. In terms of further data acquisition, both to increase the sample and to train the model, the data sources might also be improved by including more corporate textual data, such as website content, company presentations, investor relationship materials and quarterly reports. Some companies were excluded from the sample because their reports were unavailable in English. Translation would increase the sample size, although adding more reports to the sample might not improve the findings, while increasing time and costs (see Faraway and Augustin (2018)). To improve the approach and prototype and move towards further automation, ways to automate downloading of textual data might be developed. In this analysis, corporate textual data were downloaded manually. A suitable database or API might significantly improve this process and reduce the time taken to acquire data.

Finally, aviation and overall mobility are fast-paced industries influenced by large-scale global trends. The COVID-19 pandemic is currently impacting hugely on mobility companies, and the pressing issue of exogenous shocks is included as H7 in this analysis. Hence, the results might look different if the analysis were run using more recent corporate textual data, showing the limitations of the overall study design. Corporate textual data are dynamic, and trends may emerge or disappear over time, leading to a need to constantly adapt the hypotheses (classes) and classification model prototypes. This study can be considered as a starting point for such a dynamic trend-testing model. The developed dictionary (see Appendix Table 11) can be re-used with additional or updated data, and expanded with further terms or hypotheses. This paper describes the development of the supervised classification model, allowing replication with new training data.

6.2. Summary

It is important to explore the entire D2D air-travel chain in order to improve the overall journey experience and tackle passengers pain points. Trend analyses are essential for decision making, and it is as yet unclear whether managerial insights from the literature are adopted as key management topics by TSPs. This paper contributes to research and the industry by providing an innovative approach for assessing the strategic relevance of D2D air-travel trends, using prototype multi-labeled text classification models.

The study explored whether TSPs (airports, airlines and airport feeder providers) along the travel chain consider and communicate findings from travel research. Fifty-two objects of analysis (annual corporate reports and sustainability reports) were gathered and analyzed using a dictionary-based classifier model and a supervised learning model with the multinomial NB and linear SVM classifiers. Comparison of the models outputs indicates that the SVM yields the best results for answering the research question. The results reveal that TSPs attach greater relevance to environmentally-friendly air transport and related products. Disruption management, leveraging passengers data, and improving airport feeder traffic through novel mobility concepts are trends considered to have medium relevance. Establishing partnerships, personalization of travel chains and constant adaptation to exogenous shocks, such as terrorism and pandemics, are given lower relevance.

Overall, the study provides a way to overcome the problem of collecting organizational data arising from public disclosure issues. It establishes trend hypotheses and a dictionary of keywords to test the strategic relevance of current D2D air-travel trends. The dictionary-based classifier and the supervised learning model are compared empirically. In summary, this novel approach offers many opportunities for further improvement and is transferable to other research questions and contexts for trend testing.

Declaration of competing interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank Annika Paul, Lukas Preis and Damir Valput for their time and valuable input.

Appendix and supplementary material

Supplementary material associated with this article can be found in the online version. The Appendix follows below.

Crunchbase Search Protocol

Companies were searched on Crunchbase (Crunchbase, 2020) according to the following protocol:

- Founded: between 2015 and 2020
- Headquartered: Europe
- Industry (includes any): transportation, ride sharing, last mile transport, car sharing
- Operating status: active
- Exclude industries: delivery, good and beverage, customer service, health care, freight service, logistics, supply chain management, medical
- Filter: estimated revenue range (descending)
- Additional search for basic description contains "airport"
- Top 33 analyzed for purpose of this study; 11 company used in this analysis

[14] Removed from analysis as word is too ambiguous.

Table 7
The 20 largest functional urban areas of the EU, by cities and commuting zones inhabitants, data from 2017 (EUROSTAT, 2020), incl. primary public transport providers, bold names incl. in sample.

#	City	City (commuting zone) in m	Public transport provider(s)[a]
1	Paris (FR)	9,8 (12,8)	**RATP (Autonomous Parisian Transportation Administration); SNCF (French National Railway Company)**
2	London (UK)	8,8 (12,1)	**TfL (Transport for London); Arriva Rail London; Hull Trains; Grand Central railway; Merseyrail; Chiltern Railways; Virgin Trains; Virgin Trains East Coast; ScotRail; London Overground; Heathrow Connect; Eurostar**
3	Madrid (ES)	4,9 (6,6)	EMT (Madrid's Municipal Transport Corporation); RENFE (Renfe Operadora)
4	Berlin (DE)	3,6 (5,1)	BVG (Berlin Transport Company); VBB (Transport Association Berlin-Brandenburg); **DB (Deutsche Bahn)**
5	Milan (IT)	4,1 (5,1)	ATM (Azienda Trasporti Milanesi); **Trenitalia**
6	Ruhr (DE)[b]	0,6 (5,1)	**BVR (Busverkehr Rheinland)**; Vestische Straßenbahnen; Dortmunder Stadtwerke; Ruhrbahn; VRR (Transport Association Rhein-Ruhr); **DB (Deutsche Bahn)**
7	Barcelona (ES)	3,6 (4,9)	**TMB (Transports Metropolitans de Barcelona)**; RENFE (Renfe Operadora)
8	Roma (IT)	2,9 (4,4)	ATAC (Transport Company of the Municipality of Rome); **Trenitalia**
9	Naples (IT)	3,1 (3,4)	ANM (Mobility Company of Naples); **Trenitalia**
10	Greater Manchester (UK)	2,8 (3,3)	**TfGM (Transport for Greater Manchester)**; cf. London
11	Hamburg (DE)	1,8 (3,2)	HHA (Hamburger Hochbahn); HVV (Hamburg Transport Association); **DB (Deutsche Bahn)**
12	West Midlands urban area (UK)	2,5 (3,0)	**TfWM (Transport for West Midlands)**; cf. London
13	Lisbon (PT)	1,8 (2,8)	Carris (Lisbon Tramways Company); **Lisbon Metro**; CP (Trains of Portugal)
14	Budapest (HU)	1,8 (2,9)	**BKK (Centre for Budapest Transport); MAV (Hungarian State Railways)**
15	Munich (DE)	1,5 (2,8)	**MVG (Munich Transportation Company)**; MVV (Munich Transport and Tariff Association); **DB (Deutsche Bahn)**
16	Stuttgart (DE)	0,6 (2,7)	SBB (Stuttgarter Straßenbahnen AG); VVS (Transport Association Stuttgart); **DB (Deutsche Bahn)**
17	Frankfurt a. M. (DE)	0,7 (2,6)	VGF (Stadtwerke Verkehrsgesellschaft Frankfurt a. M.); RMV (Rhein-Main-Transport Association); **DB (Deutsche Bahn)**
18	Brussels (BE)	1,2 (2,6)	STIB (Brussels Intercommunal Transport Company); **NMBS / SNCB (National Railway Company of Belgium)**
19	Leeds (UK)	0,8 (2,6)	WYPTE (West Yorkshire Passenger Transport Executive); cf. London
20	Amsterdam (NL)	0,8 (2,5)	GVB (Municipal Transport Company); **NS (Nederlandse Spoorwegen)**

[a] Company names are translated into English. *Verkehrsverbunde* do not provide operational transport services

[b] Only Ruhr cities over 500.000 inhabitants considered (Kreis Recklinghausen, Dortmund, Essen)

Table 8
European airlines flying to & from Europe retrieved from OAG (2018) based on scheduled flights and planned seat capacity; 23 airlines included in sample; 30 airlines included as keywords.

#	Carrier Name (Domicile Country)	In Sample	Total Seat Capacity
1	Ryanair (IE)	x	142.540.776
2	Easyjet (UK)	x	100.082.969
3	Turkish Airlines (TR)	x	91.560.208
4	Lufthansa German Airlines (DE)	x	89.606.683
5	British Airways (UK)	x	58.891.280
6	Air France (FR)		57.403.402
7	Aeroflot Russian Airlines (RU)	x	54.704.975
8	SAS Scandinavian Airlines (SE)	x	42.433.727
9	KLM-Royal Dutch Airlines (NL)	x	39.925.170
10	Vueling Airlines (ES)	x	39.519.020
11	Eurowings (DE)	x	36.731.238
12	Wizz Air (HU)	x	36.329.808
13	Pegasus Airlines (TR)	x	33.016.116
14	Iberia (ES)	x	31.277.293
15	Alitalia - Societa Aerea Italiana (IT)		30.728.937
16	Norwegian Air Shuttle (NO)	x	25.153.111
17	Swiss (CH)	x	24.146.541
18	TAP Air Portugal (PT)	x	20.504.640
19	Norwegian (IE)	x	18.686.220
20	Finnair (FI)	x	18.561.772
21	Austrian Airlines (AT)	x	18.348.903
22	Aer Lingus (IE)	x	16.501.640
23	Siberia Airlines (RU)		15.478.859
24	Air Europa (ES)	x	14.508.791
25	Flybe (UK)		14.095.905
26	Jet2.com (UK)	x	13.573.868
27	TUI Airways (UK)		12.923.015
28	LOT - Polish Airlines (PL)		12.863.577
29	Brussels Airlines (BE)	x	12.716.949
30	Condor Flugdienst (DE)		10.166.102

Table 9
Top 30 European airports retrieved from OAG (2018) based on total planned seat capacity (departure and arrival); 22 airports included in sample; 30 airports included as keywords.

#	Airport Name (Domicile Country)	In Sample	Total Seat Capacity
1	London Heathrow Apt (UK)	x	99.805.194
2	Frankfurt Intl. Apt (DE)		88.868.689
3	Paris Charles de Gaulle Apt (FR)	x	86.530.078
4	Istanbul Ataturk Apt (TR)	x	83.343.911
5	Amsterdam Apt (NL)	x	81.387.234
6	Madrid Adolfo Suarez-Barajas Apt (ES)	x	69.021.727
7	Munich Intl. Airport (DE)		61.580.389
8	Barcelona Apt (ES)	x	60.379.660
9	Moscow Sheremetyevo Intl. Apt (RU)	x	55.283.585
10	Rome Fiumicino Apt (IT)	x	55.270.558
11	London Gatwick Apt (UK)	x	52.237.947
12	Zurich Apt (CH)	x	41.319.089
13	Paris Orly Apt (FR)	x	41.179.843
14	Istanbul Sabiha Gokcen Apt (RU)		39.638.741
15	Copenhagen Kastrup Apt (DK)	x	39.080.691
16	Oslo Gardermoen Apt (NO)	x	38.298.280
17	Vienna Intl. Apt (AT)	x	35.700.301
18	Stockholm Arlanda Apt (SE)	x	35.630.452
19	Lisbon Apt (PT)	x	34.912.145
20	Palma de Mallorca (ES)	x	33.820.858
21	Dublin Apt (IE)	x	33.294.172
22	Manchester (GB) Apt (UK)	x	32.975.110
23	Moscow Domodedovo Apt (RU)		32.585.070
24	Brussels Apt (BE)	x	31.936.415
25	London Stansted Apt (UK)	x	31.884.388
26	Duesseldorf Intl. Apt (DE)	x	31.614.106
27	Milan Malpensa Apt (IT)		31.288.309
28	Berlin Tegel Apt (DE)		29.514.664
29	Helsinki-Vantaa Apt (FI)		27.374.785
30	Athens Apt (EL)		25.393.842

Table 10
Top 40 mobility and transport-related companies; used as keywords.

#	Company Name	Short Description	H	Source
1	Sono Motors	Global mobility and energy service provider	H5	Crunchbase 2020)
2	Maas Global Oy	Mobility-as-a-Service operator	H3	Crunchbase (2020)
3	Bipi	Car subscription company	H5	Crunchbase (2020)
4	Viruo	Car Rental	H5	Crunchbase (2020)
5	Welcome Pickups	Airport pickup	H5	Crunchbase (2020)
6	Taxi2airport. com	Online booking for airport transfers	H5	Crunchbase (2020)
7	RideLink	P2P Car Rental platform (UK and Spain)	H5	Crunchbase (2020)
8	moovel Group	Urban mobility solutions	H5	Crunchbase (2020)
9	GIVT	Compensation services	H3	Crunchbase (2020)
10	BeeRides	Airport car sharing service	H5	Crunchbase (2020)
11	Blue Label	Private, low cost airport transfer	H5	Crunchbase (2020)
12	GetYourGuide	Offers tours and activities	H3	LH Inno Hub (2020b)
13	FlixBus	Intercity mobility	H5	LH Inno Hub (2020b)
14	Cabify	Ride hailing carpooling services	H5	LH Inno Hub (2020b)
15	BlaBlaCar	Ride hailing carpooling services	H5	LH Inno Hub (2020b)
16	Omio	Online search and booking	H3	LH Inno Hub (2020b)
17	Bolt	Taxi services	H5	LH Inno Hub (2020b)
18	Secret Escapes	Online travel agencies	H3	LH Inno Hub (2020b)
19	Selina	Alternative housing	H3	LH Inno Hub (2020b)
20	Cluno	Car rental & sharing	H5	LH Inno Hub (2020b)
21	Dreamlines	Online search and booking	H3	LH Inno Hub (2020b)
22	HomeToGo	Alternative housing	H3	LH Inno Hub (2020b)
23	TravelPerk	Enterprise management software (business pax)	H3	LH Inno Hub (2020b)
24	SeaBubbles	Intercity mobility	H5	LH Inno Hub (2020b)
25	Tiqets	Offers tours and activities	H3	LH Inno Hub (2020b)
26	Evaneos	Offers tours and activities	H3	LH Inno Hub (2020b)
27	Culture Trip	Digital content & inspiration	H3	LH Inno Hub (2020b)
28	C Teleport	Online travel agency	H3	LH Inno Hub (2020a)
29	Mobian	Provides seamless, singel ticketing	H3	LH Inno Hub (2020a)
30	Resolver	Compensation services	H3	LH Inno Hub (2020a)
31	LuckyTrip	Online travel agency	H3	LH Inno Hub (2020a)
32	WeSki	Online travel agency	H3	LH Inno Hub (2020a)
33	Dohop	Online travel agency	H3	LH Inno Hub (2020a)
34	car&away	Parking and renting (UK)	H5	LH Inno Hub (2020a)
35	Flightsayer (Lumo)	Flight delay predictions	H3	LH Inno Hub (2020a)
36	Total[14]	Major player renewable aviation fuels	H4	Mordor Intelligence (2020)
37	Neste	Major player renewable aviation fuels	H4	Mordor Intelligence (2020)
38	Swedish Biofuels	Major player renewable aviation fuels	H4	Mordor Intelligence (2020)
39	SG Preston	Major player renewable aviation fuels	H4	Mordor Intelligence (2020)
40	Gevo	Major player renewable aviation fuels	H4	Mordor Intelligence (2020)

Table 11
Basic dictionary, 225 entities.

Hypotheses	Metric (Keywords)
Personalization (H1)	advertis*, customis*, flexibl*, individualis*, individualiz*, MaaS, need, on-demand, pattern, passeng*, personalis*, personaliz*, prefer*, recognit*, requir*, servic*, segment*, target*
Passenger data (H2)	analyt*, app, agent, autom*, blockchain*, biometr*, board*, catbot*, data*, data-min*, digitalis*, digitaliz*, digit*, element, experi*, home-print*, ident*, ML, Machin*, mine, mix*, predict*, realiti*, robot*, self-board*, self, SITA, tag, track*, virtual*
Establishing partnerships (H3)	add-on, Amazon, AI, Artifici*, Airbnb, air-rail*, booking.com, bundl*, C Teleport, Culture Trip, co-modal, collabor*, cooper*, Dohop, Dreamlines, door-to-door*, door*, D2D*, Evaneos, Flightsayer, GIVT, GetYourGuide, Google, HomeToGo, IATA, integr*, Intellig*, interfac*, intermod*, interoper*, Lumo, LuckyTrip, Maas Global, Mobian, multimodal*, Netflix, node*, Omio, partnership*, Resolver, start-up, Skyscanner, solut*, Selina, Secret Escapes, TravelPerk, Tiqets, wemovo, WeSki, (customized keywords depending on TSP: airline, airport, railway, bus, public transport providers)
Environmentally-friendly air transport (H4)	altern*, awar*, ACA, biofuel*, carbon, CORSIA, compens*, diesel, decarbonis*, decarboniz*, drop-in, eco-friend*, energi*, environment*, environment-friend*, electrif*, emiss*, ETS, futur*, fuel, footprint, green, Gevo, Greenpeace, hydrogen*, kerosen, Neste, offset*, renew*, sustain*, shame*, SG Preston, Swedish Biofuels
Airport feed (H5)	access*, autonom*, BeeRides, Bipi, BlaBlaCar, Bolt, Blue Label, car, car&away, Cabify, concept, Cluno, driv*, egress, emerg*, FreeNow, feed*, FlixBus, hail*, innov*, last*, Lyft, mile, moovel, MyTaxi, novel, privat*, pool*, ride, RideLink, SeaBubbles, seamless*, share*, shuttl*, Sono Motors, Taxi2airport, transrapid*, Uber, Viruo, Welcome Pickups
Disruption management (H6)	automat*, alert*, assist*, cascad*, custom*, care, congest*, delay*, disturb*, disrupt*, flow, itinerari*, network, notif*, push, rebook*, re-configur*, real, real-time, status, time
Exogenous shocks (H7)	attack*, crisi*, exogen*, epidem*, hygien*, health, luggag*, measur*, monitor*, pandem*, resili*, reaction*, recov*, regul*, secur*, screen*, safeti*, shock, terror*, terrorist*, WHO

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.techfore.2021.120865.

References

accenture, 2018. Blockchain technology for airlines.
ACI EUROPE, 2009. Airport carbon accreditation.
Anandarajan, M., Hill, C., Nolan, T., 2019. Practical text analytics, 2. https://doi.org/10.1007/978-3-319-95663-3.

Bardhi, F., Eckhardt, G.M., 2012. Access-based consumption: the case of car sharing. Journal of Consumer Research 39 (4), 881–898. https://doi.org/10.1086/666376.
Bauen, A., Bitossi, N., German, L., Harris, A., Leow, K., 2020. Sustainable aviation fuels. Johnson Matthey Technol. Rev. (July).
Baumgartner, C., Kätker, J., Tura, N., 2016. Dora – integration of air transport in overall urban and regional mobility information. Transp. Res. Procedia 14, 3238–3246. https://doi.org/10.1016/j.trpro.2016.05.268.
Benoit, K., 2018. The structure of quanteda.
Benoit, K., Muhr, D., Watanabe, K., 2020. stopwords: Multilingual stopword lists.
Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., 2018. Quanteda: an r package for the quantitative analysis of textual data. Journal of Open Source Software 3 (30), 774. https://doi.org/10.21105/joss.00774.
Blalock, G., Vrinda, K., Simon, D., 2007. The impact of post-9/11 airport security measures on the demand for air travel. The Journal of Law and Economics 50 (4), 731–755.
Bloomberg Professional Services, 2018. Extracting value from social and news data.
Budd, L., Ison, S., Budd, T., 2016. Improving the environmental performance of airport surface access in the uk: the role of public transport. Research in Transportation Economics 59, 185–195. https://doi.org/10.1016/j.retrec.2016.04.013.
Crunchbase, 2020. Company search.
Das, S., Dutta, A., Brewer, M.A., 2020. Case study of trend mining in transportation research record articles. Transportation Research Record: Journal of the Transportation Research Board 3. https://doi.org/10.1177/0361198120936254.036119812093625
EcoPassenger, 2020. Compare the energy consumption, the co2 emissions and other environmental impacts for planes, cars and trains in passenger transport.
Emirates, 2019. Annual report 2018-19.
Esendemirli, E., 2014. A comparative analysis of disclosures in annual reports. Journal of Modern Accounting and Auditing 10 (4).
Esri, Peacetechlab, 2020. A map of terrorist attacks, according to wikipedia.
European Commission, 2011. Flightpath 2050 europe's vision for aviation - report of the high level group on aviation research. 10.2777/50266.
European Commission, 2019. Special eurobarometer 490 survey: Climate change report.
European Commission, 2020. Reducing emissions from aviation.
European Investment Fund, 2020. 2nd eib climate survey (2/4): Citizens' commitment to fight climate change in 2020.
Eurostat, 19.3.2020. online data codes: urb_cpop1 and urb_lpop1.
Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J., 2008. Liblinear: a library for large linear classification. Journal of machine learning research 9 (August), 1871–1874.
Faraway, J.J., Augustin, N.H., 2018. When small data beats big data. Statistics & Probability Letters 136, 142–145. https://doi.org/10.1016/j.spl.2018.02.031.
García-Albertos, P., Cantú Ros, O.G., Herranz, R., Cirruelos, C., 2017. Analyzing Door-to-door Travel Times through Mobile Phone Data. In: SESAR (Ed.), 7th SESAR Innovation Days. Belgrade, Serbia., pp. 1–8. https://doi.org/10.1007/s13272-019-00432-y.
Garikapati, V.M., Pendyala, R.M., Morris, E.A., Mokhtarian, P.L., McDonald, N., 2016. Activity patterns, time use, and travel of millenials: a generation in transition? Transport Reviews 36 (5), 558–584. https://doi.org/10.1080/01441647.2016.1197337.
Geneva Airport, 2018. Sustainability report 2016 - 2018.
Gold, L., Balal, E., Horak, T., Cheu, R.L., Mehmetoglu, T., Gurbuz, O., 2019. Health screening strategies for international air travelers during an epidemic or pandemic. Journal of Air Transport Management 75, 27–38. https://doi.org/10.1016/j.jairtraman.2018.11.006.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT Press, Cambridge, Massachusetts and London, England.
Grimmer, J., Brandon, S.M., 2013. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. Political analysis 21 (3), 267–297. https://doi.org/10.1093/pan/mps028.
Gustafson, P., 2014. Business travel from the traveller's perspective: stress, stimulation and normalization. Mobilities 9 (1), 63–83. https://doi.org/10.1080/17450101.2013.784539.
IATA, 2019. Global passenger survey - 2019 highlights.
IATA, 2020a. Guidance for cabin operations during and post pandemic.
IATA, 2020b. Nextt: let's build the journey of the future.
ICAO, 2020. Carbon offsetting and reduction scheme for international aviation (corsia).
IDG Business Media, 2020. Studie machine learning 2020.
Ivancevich, J.M., Konopaske, R., Defrank, R.S., 2003. Business travel stress: a model, propositions and managerial implications. Work & Stress 17 (2), 138–157. https://doi.org/10.1080/0267837031000153572.
Javornik, M., Nadoh, N., Lange, D., 2018. Data Is the New Oil. In: Müller, B., Meyer, G. (Eds.), Towards User-Centric Transport in Europe. Springer International Publishing, pp. 295–308.
Kadhim, A.I., 2019. Survey on supervised machine learning techniques for automatic text classification. Artif. Intell. Rev. 52 (1), 273–292. https://doi.org/10.1007/s10462-018-09677-1.
Kim, K., Park, O.J., Yun, S., Yun, H., 2017. What makes tourists feel negatively about tourism destinations? application of hybrid text mining methodology to smart destination management. Technol. Forecast. Soc. Change 123, 362–369. https://doi.org/10.1016/j.techfore.2017.01.001.
Kim, N.Y., Park, J.W., 2016. A study on the impact of airline service delays on emotional reactions and customer behavior. Journal of Air Transport Management 57, 19–25. https://doi.org/10.1016/j.jairtraman.2016.07.005.
Kinra, A., Kashi, S.B., Pereira, F.C., Combes, F., Rothengatter, W., 2019. Textual Data in Transportation Research: Techniques and Opportunities. In: Antoniou, C., Pereira, F. B., Dimitriou, L. (Eds.), Mobility patterns, big data and transport analytics. Elsevier, Amsterdam, Netherlands, pp. 173–197. https://doi.org/10.1016/B978-0-12-812970-8.00008-7.
Kluge, U., Paul, A., Ploetner, K.O., 2019. Characteristics of potential user groups of new forms of mobility using the example of urban air mobility. In: Air Transport Research Society (Ed.), 23th ATRS World Conference, pp. 1–15. https://doi.org/10.5281/zenodo.3946241.
Kluge, U., Paul, A., Urban, M., Ureta, H., 2018. Assessment of Passenger Requirements along the D2DAir Travel Chain. In: Müller, B., Meyer, G. (Eds.), Towards User-Centric Transport in Europe, Lecture Notes in Mobility. Springer International Publishing, Cham, pp. 255–276.
Kluge, U., Paul, A., Ureta, H., Ploetner, K.O., 2018. Profiling future air transport passengers in europe. In: European Commission (Ed.), 7th Transport Research Arena TRA 2018. Vienna, Austria, pp. 1–9. https://doi.org/10.5281/zenodo.1446080.
Kluge, U., Ringbeck, J., Spinler, S., 2020. Door-to-door travel in 2035 – a delphi study. Technol. Forecast. Soc. Change 157C (August 2020), 120096. https://doi.org/10.1016/j.techfore.2020.120096.
Korfiatis, N., Stamolampros, P., Kourouthanassis, P., Sagiadinos, V., 2019. Measuring service quality from unstructured data: atopic modeling application on airline passengers' online reviews. Expert Syst. Appl. 116, 472–486. https://doi.org/10.1016/j.eswa.2018.09.037.
Kowalczyk, A., 2017. Support vector machines succinctly: K. Syncfusion Inc., Morrisville, NC.
Kowsari, K., Meimandi, K.J., Heidarysafa, M., Mendu, S., Barnes, L., Brown, D., 2019. Text classification algorithms: a survey. Information 10 (4), 1–68. https://doi.org/10.3390/info10040150.
Kuhn, K.D., 2018. Using structural topic modeling to identify latent topics and trends in aviation incident reports. Transportation Research Part C: Emerging Technologies 87, 105–122. https://doi.org/10.1016/j.trc.2017.12.018.
Kumar, Y.K., Rao, V.K., 2019. Development of balanced scorecard framework for performance evaluation of airlines. International Journal of Management 10 (6), 214–234.
Landrum, N.E., Ohsowski, B., 2018. Identifying worldviews on corporate sustainability: acontent analysis of corporate sustainability reports. Business Strategy and the Environment 27 (1), 128–151. https://doi.org/10.1002/bse.1989.
Law, K.M., Breznik, K., 2018. What do airline mission statements reveal about value and strategy? Journal of Air Transport Management 70 (August 2017), 36–44. https://doi.org/10.1016/j.jairtraman.2018.04.015.
Lee, D.S., Fahey, D.W., Skowron, A., Allen, M.R., Burkhardt, U., Chen, Q., Doherty, S.J., Freeman, S., Forster, P.M., Fuglestvedt, J., Gettelman, A., de León, R.R., Lim, L.L., Lund, M.T., Millar, R.J., Owen, B., Penner, J.E., Pitari, G., Prather, M.J., Sausen, R., Wilcox, L.J., 2020. The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. Atmos. Environ. 117834. https://doi.org/10.1016/j.atmosenv.2020.117834.
Li, X., Xie, Q., Daim, T., Huang, L., 2019. Forecasting technology trends using text mining of the gaps between science and technology: the case of perovskite solar cell technology. Technol. Forecast. Soc. Change 146 (November 2017), 432–449. https://doi.org/10.1016/j.techfore.2019.01.012.
Lin, Y.H., Ryan, C., Wise, N., Low, L.W., 2018. A content analysis of airline mission statements: changing trends and contemporary components. Tourism Management Perspectives 28, 156–165. https://doi.org/10.1016/j.tmp.2018.08.005.
Linden, E., 2021. Pandemics and environmental shocks: what aviation managers should learn from covid-19 for long-term planning. Journal of Air Transport Management 90, 101944. https://doi.org/10.1016/j.jairtraman.2020.101944.
Loughran, T.I., Mcdonald, B., 2016. Textual analysis in accounting and finance: asurvey. Journal of Accounting Research 54 (4), 1187–1230. https://doi.org/10.1111/1475-679X.12123.
Lucini, F.R., Tonetto, L.M., Fogliatto, F.S., Anzanello, M.J., 2020. Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. Journal of Air Transport Management 83, 101760. https://doi.org/10.1016/j.jairtraman.2019.101760.
Lufthansa, 2019. Rail&fly.
Lufthansa Innovation Hub, 2019. The airline digital index (adix).
Lufthansa Innovation Hub, 2020a. Airlines and their startup investments.
Lufthansa Innovation Hub, 2020b. Europe's 20 top-funded travel and mobility tech startups.
Lufthansa Innovation Hub, 2020c. Über uns.
Luggage Free, 2020. The luggage shipping experts.
Manning, C.D., Raghavan, P., Schutze, H., 2008. Introduction to information retrieval. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511809071.
McKinsey & Company, 2021. How will you be getting around in 2030? it depends on where you live.
McLachlan, J., James, K., Hampson, B., 2018. Assessing whether environmental impact is a criterion of consumers when selecting an airline. International Journal of Advanced Research 6 (3), 740–755. https://doi.org/10.21474/IJAR01/6727.
Merkert, R., Bushell, J., Beck, M.J., 2020. Collaboration as a service (caas) to fully integrate public transportation – lessons from long distance travel to reimagine mobility as a service. Transportation Research Part A: Policy and Practice 131, 267–282. https://doi.org/10.1016/j.tra.2019.09.025.
Mirończuk, M.M., Protasiewicz, J., 2018. A recent overview of the state-of-the-art elements of text classification. Expert Syst. Appl. 106, 36–54. https://doi.org/10.1016/j.eswa.2018.03.058.
Monmousseau, P., Delahaye, D., Marzuoli, A., Féron, E., 2019. Door-to-door travel time analysis from paris to london and amsterdam using uber data. In: SESAR (Ed.), 9th SESAR Innovation Days, pp. 1–7.

Mordor Intelligence, 2020. Renewable aviation fuel market - growth, trends, and forecast (2020 - 2025).
OAG, 2018. Official airline guide 2018.
OAG, 2020. Oag takeoff 2020: Essential metrics on the world's major airlines.
ÖBB, 2019. Annual report 2019.
Pearce, P., 14.4.2020. 1st online panel discussionthe coronavirus outbreak: An unprecedent shock for aviation – what now?
Punel, A., Ermagun, A., 2018. Using twitter network to detect market segments in the airline industry. Journal of Air Transport Management 73, 67–76. https://doi.org/10.1016/j.jairtraman.2018.08.004.
Roberts, J.H., Kayande, U., Stremersch, S., 2014. From academic research to marketing practice: exploring the marketing science value chain. International Journal of Research in Marketing 31 (2), 127–140. https://doi.org/10.1016/j.ijresmar.2013.07.006.
Rochwerger, A. S., 2019. A better approach to ai: Realize business value first.
Rose, J., Lennerholt, C., 2017. Low cost text mining as a strategy for qualitative researchers. Electronic Journal of Business Research Methods 15 (1), 2–16.
Rothfeld, R., Straubinger, A., Paul, A., Antoniou, C., 2019. Analysis of european airports' access and egress travel times using google maps. Transp. Policy (Oxf) 81, 148–162. https://doi.org/10.1016/j.tranpol.2019.05.021.
Schmitt, D., Gollnick, V., 2016. The Air Transport System. In: Schmitt, D., Gollnick, V. (Eds.), Air Transport System. Springer Vienna, Vienna, pp. 1–17. https://doi.org/10.1007/978-3-7091-1880-1_1.
Schulz, T., Gewald, H., Böhm, M., 2018. The long and winding road to smart integration of door-to-door mobility services: an analysis of the hindering influence of intra-role conflicts. Twenty-Sixth European Conference on Information Systems (ECIS).
Seo, Gang-Hoon, 2020. A Content Analysis of International Airline Alliances Mission Statements. Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy 11 (1), 89–105. https://doi.org/10.2478/bsrj-2020-0007.
Sezgen, E., Mason, K.J., Mayer, R., 2019. Voice of airline passenger: a text mining approach to understand customer satisfaction. Journal of Air Transport Management 77, 65–74.
Silge, J., Robinson, D., 2017. Text mining with r. O'Reilly Media, USA.
Siren, A., Haustein, S., 2013. Baby boomers ' mobility patterns and preferences : what are the implications for future transport ? Transp. Policy (Oxf) 29, 136–144.
Siren, A., Haustein, S., 2015. Older people's mobility segments factors trends. Transport Reviews 35 (4), 466–487.
SITA, 2019. 2025: Air travel for a digital age (white paper).
Škare, M., Soriano, D.R., Porada-Rochoń, M., 2020. Impact of covid-19 on the travel and tourism industry. Technological Forecasting & Social Change 120469. https://doi.org/10.1016/j.techfore.2020.120469.
Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information Processing & Management 45 (4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.
Sun, X., Wandelt, S., Stumpf, E., 2018. Competitiveness of on-demand air taxis regarding door-to-door travel time: a race through europe. Transportation Research Part E: Logistics and Transportation Review 119, 1–18. https://doi.org/10.1016/j.tre.2018.09.006.
Susilo, Y.O., oodcock, A.W., Liotopoulos, F., Duarte, A., Osmond, J., Abenoza, R.F., nghel, L.E.A., Herrero, D., Fornari, F., Tolio, V., O'Connell, E., Markucevič i?tė, I., Kritharioti, C., Pirra, M., 2017. Deploying traditional and smartphone app survey methods in measuring door-to-door travel satisfaction in eight european cities. Transp. Res. Procedia 25, 2257–2275.
Ting, S.L., Ip, W.H., Tsang, A.H.C., 2011. Is naive bayes a good classifier for document classification. International Journal of Software Engineering and Its Applications 5 (3), 37–46.
TSA, 30.12.2002. Tsa meeting december 31, deadline for screening all checked baggage: Press release.
Ureta, H., Christobal, S., Belkoura, S., Gómez, I., Cook, A. J., Tanner, G., Gurtner, G., Paul, A., Kluge, U., Hullah, P., 2017. Assessment Execution: DATASET2050 D5. 2. Madrid, Spain.
Vapnik, V.N., 1995. The nature of statistical learning theory. Springer, New York. https://doi.org/10.1007/978-1-4757-3264-1.
Wade, B., Topalova, Y., Boutin, N., Jhunjhunwala, P., Loh, H.-H., von Oertzen, T., Ukon, M., Wise, A., 2020. Seven trends that will reshape the airline industry.
Watanabe, K., Zhou, Y., 2020. Theory-driven analysis of large corpora: semisupervised topic classification of the un speeches. Soc. Sci. Comput. Rev. 3 https://doi.org/10.1177/0894439320907027.089443932090702
Welbers, K., van Atteveldt, W., Benoit, K., 2017. Text analysis in r. Commun. Methods Meas. 11 (4), 245–265. https://doi.org/10.1080/19312458.2017.1387238.
Wenzel, M., Stanske, S., Lieberman, M.B., 2020. Strategic responses to crisis. Strategic Management Journal. https://doi.org/10.1002/smj.3161.
Young, M., Farber, S., 2019. The who, why, and when of uber and other ride-hailing trips : an examination of a large sample household travel survey. Transportation Research Part A: Policy and Practice 119 (October 2018), 383–392. https://doi.org/10.1016/j.tra.2018.11.018.

Ulrike Schmalz is a researcher at Bauhaus Luftfahrt in Munich, an interdisciplinary aviation and mobility research institute looking into long-term developments. She is part of the operations team and conducts various EU- and industry projects focusing on future trends within the overall mobility sector, current and future passenger requirements towards the transport system and intermodal, door-to-door travel. Prior to her engagement at Bauhaus Luftfahrt, she gained experience working as a consultant from 2014 and 2016. She studied corporate management, economic, and social psychology in Germany and England. Since 2018, she is also an external doctoral student at WHU - Otto-Beisheim-School of Management.

Jürgen Ringbeck, former Senior Partner with Booz Allen Hamilton / Booz & Company, is working as an independent strategic investor and consultant since spring 2014. After lecturing at the WHU since 2012, he was appointed honorary professor by WHU in 2014. Professor Ringbeck held lectures on various topics of corporate management and currently holds a lecture on Transportation Management in the Master of Science Program.

Stefan Spinler holds the Chair of Logistics Management at the WHU. His research has been presented at international conferences and leading US business schools and was awarded a number of prizes, most notably the Management Science Strategic Innovation Award (from EURO) and the GOR dissertation award. Upon the completion of his doctoral studies, Stefan spent a year lecturing at the Wharton School. He has also been invited to teach at MIT as a guest professor. His postdoctoral degree (Habilitation) covered aspects of market-based supply chain coordination and was completed in 2008.