Applications of Predictive Analytics: Evolution, Integration in Modern Society, and Ethical Concerns

Trevor J. Schwecke

Department of Actuarial Science, University of Wisconsin-Milwaukee

Abstract

In a world that's changing at an unprecedented rate, the need for implementation of predictive analytics is becoming necessary across industries. With the rapid development of new technology, access to big data in order to predict future trends is easily accessible. In doing this, algorithms and models are implemented to improve accuracy. This paper discusses how these topics apply to a variety of career fields, and how the process derives from a competitiveness in market advantages, assisted by business intelligence. Further, it will discuss how the Society of Actuaries (SOA) has integrated an exam specifically relating to predictive analytics and how the UW-Milwaukee and Marquette communities have shifted their focus towards predictive analytics. Some of the most controversial and challenging matters with predictive analytics are the limitations, such as the ability to gather clean data, affordability of technology, and limited expertise in the field. In contrast, the benefits of predictive analytics include saving resources, lowering cost and mitigating risk. There are ethical concerns that come with the infatuation of big data in a market economy as companies go to extremes to gather such data; this paper proposes a more ethical method to protect the privacy of consumers, specifically relating to social media outlets. This paper concludes with forecasting how predictive analytics will look in the future, as well as the challenges that come with it.

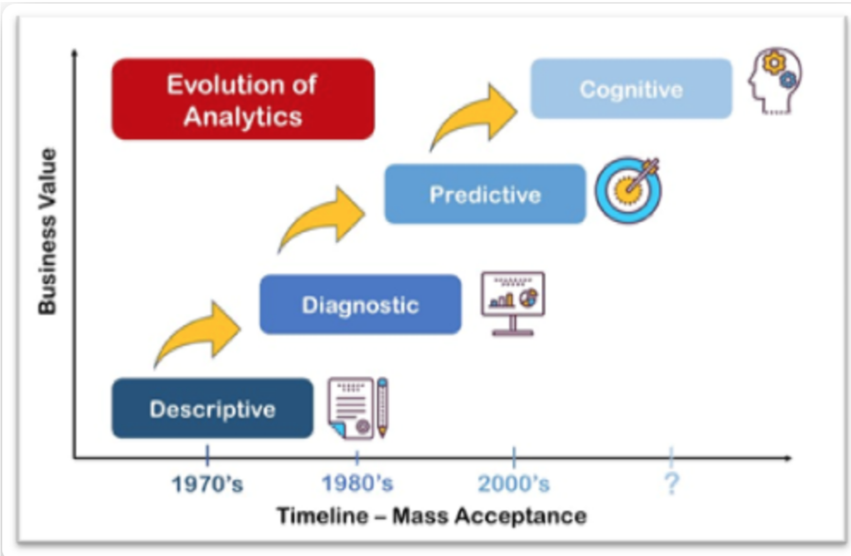
Keywords: predictive analytics, business intelligence, big data, algorithms, artificial intelligence, ethical concerns

Applications of Predictive Analytics: Evolution, Integration in Modern Society, and Ethical Concerns

The most beneficial way to gather an understanding of predictive analytics is to appreciate its' origin. The first documented use of predictive analytics was with the Lloyd Company in 1689, utilized in the underwriting process of insurance for sea voyages (Predictive Success Corporation, 2019). Even with limited data, Lloyd used data sets of prior trips to assess risk and predict liability patterns. What followed centuries later on began the descriptive phase, powered by the invention of the computer; the introduction of technology allowed for the creation of models and algorithms to forecast the future given historical data sets. Next came the diagnostic phase which concentrated on answering the question of "why". There was a vast improvement in computer functions and storage, which made possible the storage and analysis of larger data sets. The predictive phase surpassed the diagnostic phase, which is characterized as "the development of mathematical algorithms using weightings and scores to make predictions about the future…utilizing techniques from data mining, statistics, modeling, machine learning, and artificial intelligence" (Adelsberg, 2018). During this phase, the Predictive Index was created in order to assess the qualifications of applicants for the job position applied for; these are based on weighted cognitive and behavioral scores. This system is utilized in a variety of career fields and is seen as one of the earliest developments derived from predictive analytics. The final and current stage of predictive analytics is the cognitive phase; this phase accounts for rapidly changing technology, issuing the need for machine learning; a proper understanding of artificial intelligence allows for a more efficient and cost-effective process. Figure 1 illustrates and contextualizes the evolution of predictive analytics.

Figure 1

Evolution of Predictive Analytics



Note. Figure was constructed by After, Inc. 2018.

Literature Review

Methods of Predictive Analytics

While the process is incorporated differently by career field, there are seven essential steps in the execution of predictive analytics. Throughout this process, there are three crucial components to the success of the structure. These three components include data, statistics, and assumptions (Chandrasekhar, 2019). Large volumes of data that aren't comprehendible through traditional modeling techniques, known as big data, call for a more modernized approach. Software and artificial intelligence are used to aid in the process of organizing data. Statistics, which are analyzed through technology, are predicted for an unknown variable based off of the data from one or more known parameters. Finally, assumptions on the predictors are crucial given that 100% accuracy of the future can never be obtained. These three components are crucial to the predictive analytics process, which is as follows:

Project Definition: This is a structured outline of inputs that will be used as well as anticipated outcomes.

Data Collection: Research is required to gather large volume of raw data in efforts to improve accuracy. These will be typically be organized into a data lake which "contains information in a raw state" (Brooke, 2018).

Data Analysis: This is arguably the most labor-intensive step, entailing cleaning and organizing data. Once a structured format is set for the data, it is time to run an analysis, which is can be done through regression models

Statistics: Hypothesis testing is crucial in this stage; these are typically done to show correlation between one/multiple parameters and a dependent variable.

Modeling: Computer programs, such as R studio, are utilized to model trends or patterns in the data. Human intervention is crucial in modeling, as there needs to be a clear understanding of the software.

Deployment: Once the model is finalized, it will need to be assimilated into the workplace for its designated field. This gives factual evidence for business decisions, rather than making decisions based off of intuition.

Monitoring: Since consumer behavior and modern trends are constantly changing, monitoring the model and making adjustments is just as important as the application of the model.

In summary, the predictive analytics process can be effective if each step is executed properly. The involvement of big data provides more accuracy and mitigates risk but may force companies to sacrifice cost and time in order to generate models. As a result, predictive analytics is used widespread, but sectors may incorporate it in different ways.

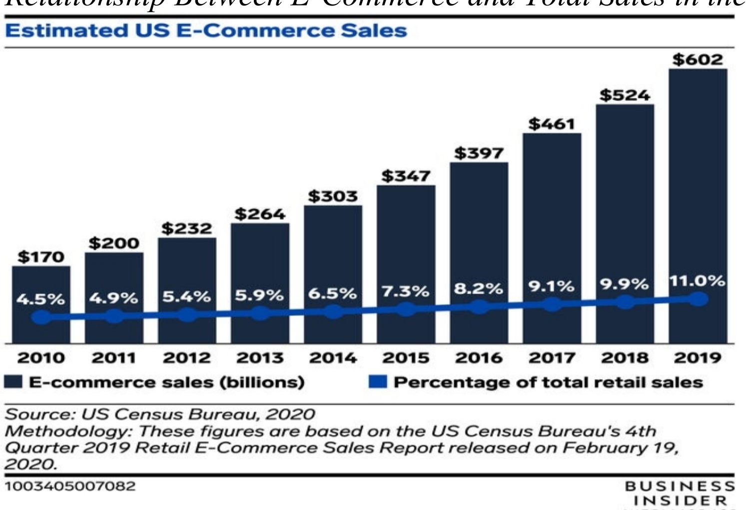
Integration of Predictive Analytics in Varying Sectors

With a focus on predicting consumer behavior and minimizing future risks, predictive analytics is seen in a multitude of career fields. Even though the process differentiates, the objective remains the same: identify futures outcomes based off historical data and trends. With the aid of artificial intelligence and skilled professionals, each sector attempts to successfully predict the future.

E-Commerce and Retail. Arguably, the sector that can most easily access consumer behavior patterns and, as a result, is dependent upon predictive analytics the most is the retail sector. Even more specifically, an area that has the most evident growth in sales numbers is the e-commerce sector. Easy accessibility and time-efficiency of online shopping coupled with the rapid technological growth demonstrates the reasoning that consumers are more inclined to purchase online. As seen by Figure 2, E-commerce sales accounted for 11% of total retail sales in 2019, along with a 14.8% growth from 2018 to 2019. This shift in demand makes it that much easier to forecast consumer behavior.

Figure 2

Relationship Between E-Commerce and Total Sales in the Last Decade



Note. Graph was taken from Business Insider Intelligence, 2020.

Companies like Amazon and Alibaba have the convenience of recording all sales into a database and utilizing the process of predictive analytics to fluctuate supply at different locations of specific goods or set competitive price levels. Still, all retail sales need to account for multiple variables that can cause a shift in consumer behavior, such as seasonal shift, economic changes, and updates with modern trend styles.

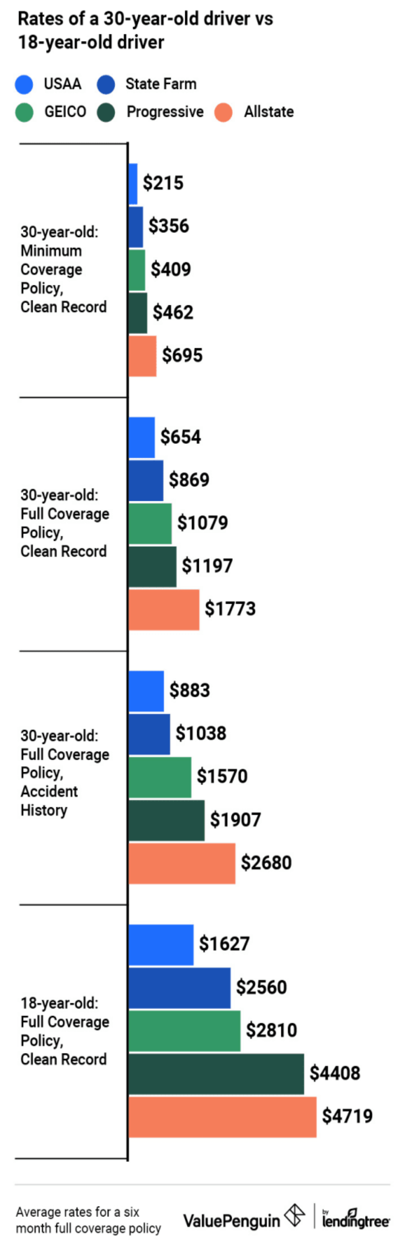
Healthcare/Medical Records. With over 36 million patients hospitalized at any given moment, there is clear indication that the health sector has an insurmountable amount of data (Pickell, 2019). Even though there is no clear indicative variable toward future illnesses or pandemics, as seen by COVID-19, the healthcare sector primarily uses predictive analytics to reduce individualized health concerns. For example, Heath Catalyst is a company that has successfully used datasets to “detect the likelihood of patients susceptible to central-line associated bloodstream infections” (Bharadwaj, 2019). In doing so, hospitals are better prepared and can monitor symptoms more closely in that demographic of people. In addition, genetic modeling is useful to determine the probability that a patient can develop health conditions that may be present in their family. In correspondence, insurance companies use this data to determine rates for different policies.

Climate/Weather. Given the definition of “forecast”, meteorologists deploy predictive analytics to help with estimating weather patterns. Contrary to popular belief, weather forecasts have been seemingly more accurate than prior years. Through the use of satellite imagery and historical data, meteorologists are able to predict forecasts for up to 30 days in advance (Pickell, 2019). Satellite imagery can help prepare communities for natural disasters that are projected to go through their area, such as hurricanes along the southeastern coast of the U.S.

Insurance. Insurance companies are constantly trying to set competitive prices for insurance policies, while trying to maximize profits. As imagined, insurance rates vary based off numerous variables; risk assessment strategies attempt to mitigate risk while providing an affordable policy premium. Figure 3 illustrates the differences in car insurance pricing by company based off three important variables: age, driving record, and type of coverage. Extremities in all three variables will result in a higher premium than an experienced driver with a clean record looking for minimum coverage. This risk tolerance model provides security and liability to auto-insurance companies to be able to cover losses as a result of a car accident. Health and life insurance policies use a similar strategy to provide coverage for their clients, while also covering their liabilities. One important contrast between strategies employed by life insurance companies versus car insurance companies is the amount of time a policy is valid for. Because an average car insurance policy is written for a shorter time frame than health/life insurance policy, companies need to consider which coverage is exercised more frequently; to no surprise, car accidents are more likely to occur than life-threatening events. Insurance companies are consistently using predictive analytics in their pricing models to maximize profit and mitigate risk.

Figure 3

Car Insurance Rates Based on Various Parameters



Note. This graph shows average rates for a six-month full coverage policy taken from ValuePenguin, 2020.

Sports. Data analytics teams within each organization are always in the search for new ways to gain a competitive edge over their opponents. The sport that has incorporated statistical models into their practice the most is baseball. Two specific categories that forecast future trends are player valuation and player development. Player valuation utilizes historical data of a player's career to determine his valuation to an organization; this is prevalent through the arbitration process, where a player argues his case for his contract valuation against an organization. In terms of player development, innovative technology and statistical models account for the training methods of players. For example, Blast Motion is a company that produces a sensor that detects various pre-contact metrics for hitters; Rapsodo is a company that produces technology to detect post-contact metrics. This allows organizations to develop multi-linear regression models to illustrate if a relationship exists between pre-contact parameters and post-contact results, such as the relationship between bat speed, attack angle and distance traveled. With the increased focus on data analytics in the sports sector, player development with drastically shift towards a dependency on predictive analytics.

Educational Impacts in the Milwaukee Community

The growth in predictive analytics has led to a change in academic study, research, and community interest. The UWM School of Continuing Education offers a certificate for Data Analysis that stresses the principles of predictive analytics. According to Bollander (2018), these include:

- Assessing the validity and quality of data types
- Translation of complex data and systems
- Use data to create predictive models
- Generate clear visual representations of the data

- Identify privacy concerns for data collection and handling

In addition, UW-Milwaukee is offering a separate course taught by Corey Fritsch, who is an experienced computer technician and data assessment coordinator. This course is a two-part course that dives into each step of the predictive analysis process, while providing real-life scenarios where each phase is used. At the notable neighboring college system, Marquette University offers a variety of undergraduate and graduate programs that emphasize the importance of data analytics, ranging from Applied Statistics to Sports and Exercise Data Analytics. Marquette University and UW-Milwaukee have partnered with Northwestern Mutual Data Science Institute to "inspire and cultivate passion for data science in the Milwaukee region" (Marquette University, 2020). This recent partnership has provided opportunities for students to conduct independent research with modernized technology and experienced industries. For aspiring actuaries, the SOA has adopted a new exam to test the problem-solving capability of candidates in respect to data analysis. As of 2018, Predictive Analytics (PA) exam is offered after the completion of the Statistics for Risk Modeling (SRM) exam (Society of Actuaries, 2020). The primary learning objectives of the exam follow the experimental design of predictive analytics, stressing the importance of being able to identify a problem, gather data and design a model, and communicate results from the trial. The exam uses a variety of software to design modules, including R studio, Microsoft Excel, and Microsoft Word to assist in analyzing big data.

Machine Learning and Artificial Intelligence

Given the enormous amount of data that industries have access to, artificial intelligence is inherently crucial to the predictive analytics process. Artificial intelligence, or machine learning, has two primary focuses with big data: supervised and unsupervised learning. Under the category of supervised learning, there is “ground truth”, or prior knowledge and expected values for what the output should be (Soni, 2018). Classification and regression are standard procedures of data analysis that fall under the umbrella of supervised learning. Classification is self-explanatory: grouping the output into classes. If the data is grouped into two classes, it is known as binary classification; if there are more than two classes involved, it is known as multiclass classification. An example of classification would be predicting if a patient could develop asthma later in their life. Regression models are used to predict output values within a range given an input. As a result, this involves more quantitative analysis. A common example is a GPS; the data system is analyzing a variety of current variables (weather or traffic) and culminates that with past trips to calculate an expected arrival time. With both classification and regression, the goal is to develop a model and determine relationships in the data to give us an expected result that is consistent with prior knowledge. Dissimilarly, unsupervised learning is done to extract patterns within a data set through clustering or association. Because there is no prior knowledge used in the process of unsupervised learning, the primary focus is to categorize the data. For example, clustering is a method that discovers patterns and structures within a data set; once a new data entry is entered, it is placed in one of the clusters that it most closely resembles. Association rules are utilized to not only illustrate commonalities amongst data but analyze and forecast trends within the data set. Association could create relationships between two subsets of different categories. For example, a fast food restaurant near a university could collect data on the students and use clustering to separate the data by two categories: age and number of days spent eating out. If restaurants discover a relationship that exists that younger college students spend more time eating out, marketing teams could develop strategies to target that demographic.

Through machine learning, artificial intelligence has proven to be more cost-effective, time efficient, and has revolutionized the way that data analytics is used. In the healthcare sector, the number of patient monitoring devices, used to create AI models, is projected to increase from 31,000 in 2017 to 3.1 million by 2021 in the United States and Israel (ABI Research, 2018). The ability to conduct more research, produce more effective drugs, and be able to forecast pandemics have all resulted from this shift in focus towards predictive analytics by the health industry. As well, rapid results from artificial intelligence allow for a more efficient use of time by healthcare workers and available treatment for patients. Furthermore, hospitals in the United States are projected to save as much as 52 billion dollars by 2021 (ABI Research, 2018).

One of the most prevalent changes that predictive analytics has shown is the emphasis on numerical evidence versus intuition. Predictive analytics, even though it may not be the most simplistic, has shown to be the most efficient process in the decision-making process. History has shown that humans have an innate thought process to make decisions weighted more on emotion than logic. Predictive analytics has given people a new perspective to consider on what the data believes to be the most logical choice rather than what intuition says.

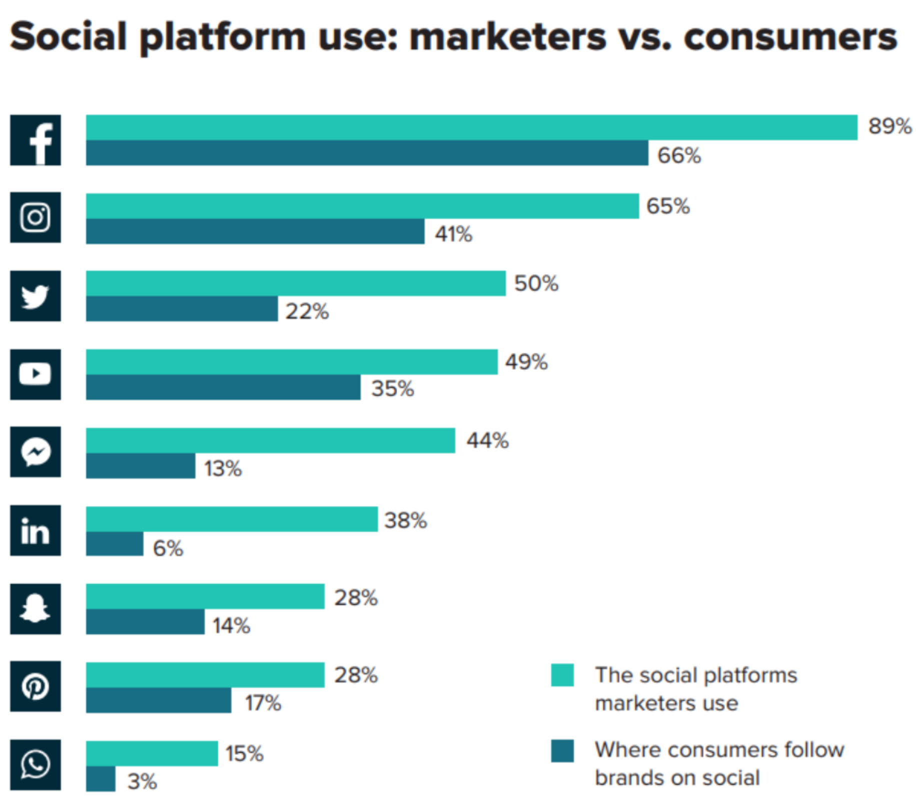
Ethical Dilemmas with User Privacy

The extents by which companies go to gather big data has drawn attention; specifically, the tactics displayed by social media companies in *The Social Dilemma* has brought forth a shocking display of manipulation. Social media engineers have created services that track users' likes and distastes, such as the "retweet" option on Twitter; the notion that social media companies are tracking a users' activity shows an infringement of privacy. Twitter is a particular platform that has seen exposure, which is regarded as a popular platform for expressing beliefs and is viewed as a prominent news outlet. In the film, Tristan Harris, a Google Design Ethicist, expresses how Twitter collects data through controls such as "like" and "retweet" to alter what a user's feed consists of (Rhodes, 2020, 0:50:01). This is especially seen with politics, as users are more likely to be shown a news source or article aligning with their political ideology to increase interaction. On Instagram, the "Terms and Conditions" agreement, which is commonly overlooked, gives access to use the microphone on users' smartphones or tablets at all times, allowing the company to invade users' privacy. Outside of being able to detect what is spoken, social media companies also gather data such as amount of time spent per post, time of day where users' interaction is the highest, and commonalities amongst shared posts (Rhodes, 2020). In the movie, Rhodes exaggerates the power that social media companies have through different scenarios; one shocking example is tracking users' movements to strategically send notifications during periods of boredom. Even though these companies don't have the resources (as of yet) to be able to do this, it provides insight that social media companies have access to more personal data than what people think. The disregard for privacy by social media companies has raised awareness by the American people, revolving around the question "How much of what American people see on social media is a result of free will?'.

The driving force behind this strategy is the same element that humans possess greed for: monetary gain. Not only is this data sold to companies to increase the revenue stream from advertisements, but an internal market exists between tech companies. There are, however, regulation parameters on purchasing and selling of data through the Fair Credit Reporting Act; in addition, some states have placed restrictions on data brokers, which are companies that collect data or buy from other companies (Matsakis, 2019). Due to limited resources and political schematics, these laws typically go unregulated. Because of this, a market exists between social media and web browser companies. Google, for instance, will sell data from their search engine to social media companies, like Instagram, to display advertisements that are tailored to each user's interests. As a part of any marketing strategy, businesses will pay these companies to display advertisements for their products or services. The marketing sector of social media companies delegate a team to develop models that efficiently advertises products or services to the proper audience. Figure 4 illustrates the relationship that exists between usage by marketers and consumers; it is no surprise that the platforms where users closely follow brands are also the platforms most used by marketers.

Figure 4

Platform Usage for Marketers vs. Consumers



Note. This illustration shows social media usage and was taken from Sprout Social, 2020.

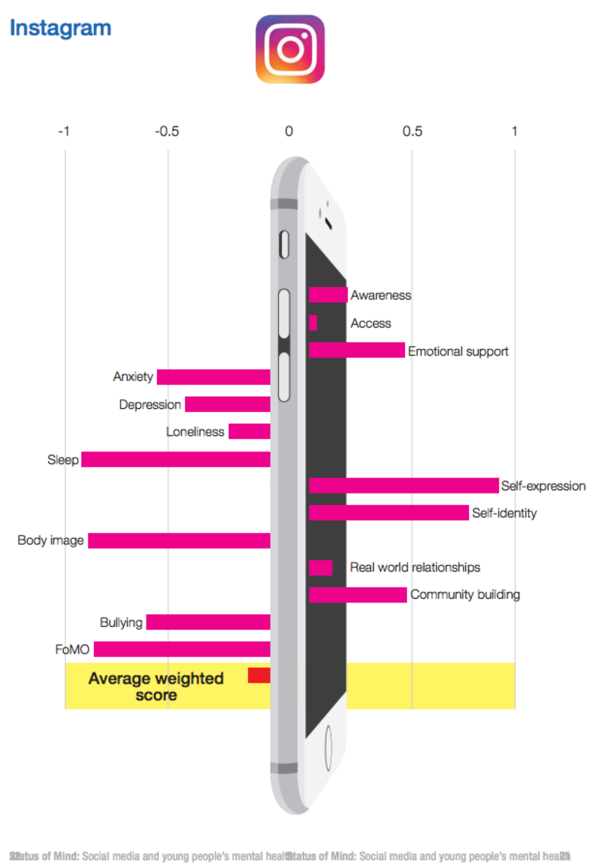
Ironically, the platform designed by Mark Zuckerberg which had intentions of keeping people up to date with close relatives and friends transitioned into the leading platform used by marketers (and by a wide margin). In conclusion, a triangular market exists between social media companies, web browsing companies, and businesses to provide individualized advertisements in hopes of driving sales or gaining attention.

Psychological Manipulation Through Data Mining

There's no denying teenagers use social media as a means of happiness through the approval of others, but who is responsible for normalizing this notion? By attempting to control the attention span of the most vulnerable age group, social media companies have created a nationwide addiction to social media apps. As previously stated, this is done through the data mining process as well as the ability to send push notifications to users based on numerous variables. This, however, has led to unforeseen complications: mental illnesses. The U.S. suicide rate for girls aged 15-19 has grown 70% (relative to the average from 2001-2010) since 2010; this, ironically, was when social media was introduced in the United States (Rhodes, 2020). Even more shockingly, the suicide rate for girls aged 10-14 has risen over 150% since the introduction of social media (Rhodes, 2020, 0:41:04). So, what is attributable to this statistic? As previously discussed, social media companies will display content to users that will gather attention and have the longest views of screen time. Thus, by consistently displaying ideas of perfection without exposing reality, teenagers are consumed by the idea of what normality truly is. For instance, a teenage girl that spends more time viewing the profile of a model is going to be shown similar content on that application. Since the teenage years are crucial to the development of the prefrontal cortex, which is responsible for the decision-making process, the consistent exposure to a near perfect body results in obsession with body image, leading to further mental complications. Instagram has been one of the fastest growing networks, rising from approximately 90 million monthly users in January of 2013 to 1 billion monthly users in June of 2018 (Chenn, 2020). Figure 5 shows the

Figure 5

Weighted Psychological and Physical Effects of Instagram Use



Note. This illustration was constructed by Lena Firestone at PsychAlive, 2017.

average weighted scores of the positive and negative effects social media has had on personal health categories; this study was conducted on 1,500 people between the ages of 14 and 24. Because Instagram's primary focus is the sharing of personal photos, it is no surprise that body image has one of the most adverse effects of Instagram's intentions. In addition, all of the negative effects have a sort form of relationship to one another, either causal or correlated. However, it's arguable that social media companies should not be held entirely responsible for this shocking statistic, as this was not their original intention. An application that was originally intended to share experiences through documented photos has turned into a community that glorifies false senses of perfection and happiness.

Solutions to Ethical Concerns

The exposure to inhumane methods of social media companies has spread awareness but is inconclusive on how to resolve the issue. One would imagine the software engineers responsible for creating the algorithms to retrieve data would also have the power to find a solution. Even though there has been an attempt to reduce the amount of information that can be retrieved by social media companies, these regulations have shown the be corrupt. Government intervention has been recommended to strengthen the laws and regulations set in place, but is that truly enough power to overthrow the power and alliance of big tech companies? The final option lays the power into the group that has fallen victim: the users. The only way to guarantee protection for users is to change the structure of ownership. Collier (2020) proposes a new model that follows suit of a method called "platform cooperativism", which aims to transfer ownership from corporations to the members. Not only does this shift in ownership provide power to the users, it grants privacy and control of the data. Also, changes in all regulations would be voted upon, similar to a democracy, by the users. Even though there are problems that would arise due to the enormous number of members, this approach seems to reduce the possibility of corporations monopolizing the social media market. A more unorthodox approach is the elimination of social media as a whole; while this idea seems far-fetched, it would eliminate those that fall victim to the malicious strategies employed by tech companies. While there was a brief reduction in the number of users on primary sources of social media, such as Twitter and Instagram, following the release of *The Social Dilemma*, these changes seem to be only temporary. The addiction and dependence on social media by Americans have deeper roots which would require further psychological research to completely comprehend.

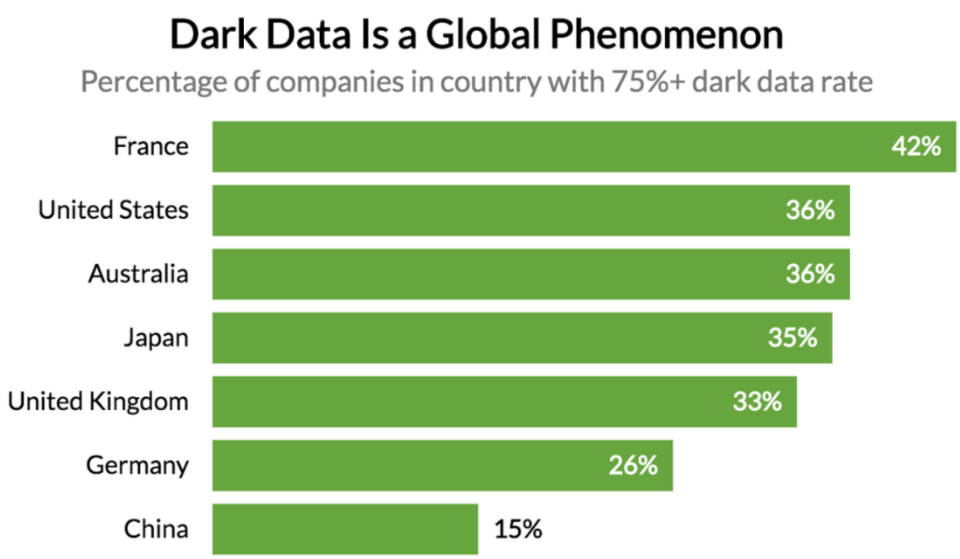
Challenges Facing Predictive Analytics

It seems a bit ironic to forecast the complications that will arise from the predictive analytics process; however, predictive analytics is a rapidly changing process that needs to adapt to technological changes, account for the introduction of marketing tools, and provide up to date training. Continuous updates to technology require capital, resources, and manpower. Software and applications are also constantly being updated, requiring businesses to both keep updated technology and staff training to understand and interpret what models are expressing. A recent 2020 study showed that executives showed an average rating of 55.8 (on a scale of 1-100) in regard to digital IQ and 25% of executives believed digital transformation would never be entirely completed (McCann, 2020). This can be attributed to the fact that technology is changing at such a rapid pace that companies cannot stay up to date and consistent staff training in data analytics is essential. Staff training and constant adaptations to changes in technology are the biggest hurdles businesses will face in the future for predictive analytics.

As previously discussed, access to data is often the least of worries for businesses; conversely, validating the data is another discussion. Because consumer behavior is constantly changing, data is quickly outdated or does not provide an accurate description of consumer behavior. As seen through political campaigns and advertisements, businesses and politicians tend to “bend the truth” in order to gain more attention from their target audience. One method data can be hidden is through the classification of data as "dark data" (Priceonomics Data Studio, 2019). This is defined as "data that has been collected, but is unstructured and, therefore, not currently being used… it is data that has been continuously collected and stored, but has not been organized via categorization, labels, or any other effective organizational tool' (Platts 2018). Ultimately, businesses have the ability to sort through data to effectively choose the quantity and quality of data that will better help their cause. Figure 6 shows how prevalent dark data is within the majority of companies in developed countries, with the United States having the second largest percentage of companies with at least 75% dark data rate.

Figure 6

Dark Data Rates in Developed Countries



Note. This graphic was taken from Splunk and True Global Intelligence, 2019.

As expected, this number will continually rise as the rate at with data is mined will only increase. Because the majority of data is unstructured, companies would have to utilize advanced business tools, costing more time, money, and resources. One possible solution to this would be the introduction of a new data system that has more parameters and structure to its storage to save time and money. Companies will continue to battle against everchanging trends and technological advancements unless there is an adequate amount of resources available or external regulation is established universally.

Conclusion

Predictive analytics has become growingly apparent in the last century. Algorithms and models are rapidly changing due to shifting consumer behavior; thus, the process defined by predictive analytics is constantly being updated within each respective sector. Predictive analytics can be most prevalently seen in the health and insurance division but is making a breakthrough in sports and social media. As a result, infringement of privacy through data mining by social media companies has brought awareness to the unethical practices by founders, such as Mark Zuckerberg. Even though the power of modern technology is stunning, predictive analytics can and should be improved through a cleaner, more ethical, and cost-effective process to enhance the accuracy of its' results.

References

ABI Research. (2018, June 4). *AI to Save Healthcare Sector US$52 Billion in 2021*. Cision PR Newswire. https://www.prnewswire.com/news-releases/ai-to-save-healthcare-sector-us52-billion-in-2021-300659060.html

Adelsberg, D. (2018, December 28). *A Brief History of Predictive Analytics Part 1*. AfterInc. https://afterinc.com/2018/12/28/brief-history-predictive-analytics-part-1/

Bollander, A. (2020, September 18). *Data Analysis Certificate*. School of Continuing Education. https://uwm.edu/sce/certificates/data-analysis-certificate/

Brooke, S. (2018, September 12). *Seven Stages of Predictive Analytics...* Raconteur. https://www.raconteur.net/technology/seven-stages-of-predictive-analytics-implementation

Chandrasekhar, S. (2019, January 11). *What is Predictive Analytics? 7 Stages of Predictive Analytics Explained*. 1redDrop. https://1reddrop.com/2019/01/11/what-is-predictive-analytics-7-stages-of-predictive-analytics-explained/

Chenn, J. (2020, August 5). *Important Instagram stats that you need to know for 2020*. SproutSocial. https://sproutsocial.com/insights/instagram-stats/#ig-user

Collier, A. (2020, October 22). *Beyond "The Social Dilemma" to social solutions*. NetFamilyNews.Org. https://www.netfamilynews.org/beyond-the-social-dilemma-to-social-solutions

Davenport, T. (2016, June 21). *A Predictive Analytics Primer*. Harvard Business Review. https://hbr.org/2014/09/a-predictive-analytics-primer

Gordon, C. (2018, December 12). *Blog #1: The History and Roots of Predictive Analytics*. SalesChoice | An AI Solution for B2B Sales. https://www.saleschoice.com/the-history-and-roots-of-predictive-analytics/

Kaput, M. (2020, January 15). *AI for Predictive Analytics: Everything You Need to Know*. Marketing Artificial Intelligence Institute. https://www.marketingaiinstitute.com/blog/ai-for-predictive-analytics

Keyes, D. (2020, February 24). *E-commerce Sales*. Business Insider. https://www.businessinsider.com/ecommerce-topped-10-percent-of-us-retail-in-2019-2020-2?international=true&r=US&IR=T

Marquette University. (2020). *Partnerships // Data Science at Marquette // Marquette University*. https://www.marquette.edu/data-science/marquette-partnerships.php

Matsakis, L. (2019, February 19). *Personal Data Collection: The Complete WIRED Guide*. Wired. https://www.wired.com/story/wired-guide-personal-data-collection/

McCann, D. (2020, April 23). *Digital Transformation: Defying Definition*. CFO. https://www.cfo.com/technology/2020/04/defying-definition/

Parthasarathy, S. (2020, September 16). *Top 5 Predictive Analytics Models and Algorithms | Logi Analytics Blog*. Logi Analytics. https://www.logianalytics.com/predictive-analytics/predictive-algorithms-and-models/

Patrizio, A. (2019, March 21). *Top Predictive Analytics Examples: Analytics for Business Success*. Datamation. https://www.datamation.com/big-data/predictive-analytics-examples.html

Pickell, D. (2019, May 14). *8 Examples of Industries Using Predictive Analytics Today*. Learning Hub. https://learn.g2.com/predictive-analytics-examples

Platts, J. (2018, June 15). *What is Dark Data? - Jill Platts*. Medium. https://medium.com/@jillplatts/what-is-dark-data-1beef317bb2e

Predictive Success Corporation. (2019, May 9). *A Brief History of Predictive Analytics - Predictive Success Corporation*. Medium. https://medium.com/@predictivesuccess/a-brief-history-of-predictive-analytics-f05a9e55145f

Priceonomics Data Studio. (2019, August 19). *Companies Collect a Lot of Data, But How Much Do They Actually Use?* Priceonomics. https://priceonomics.com/companies-collect-a-lot-of-data-but-how-much-do/

Rhodes, L. (Producer) & Orlowski, J. (Director). (2020). *The Social Dilemma* [Documentary]. United States of America. Exposure Labs Argent Pictures

Singh, A. (2019, August 21). *How AI is Improving Predictive Analytics*. MarTech Vibe. https://www.martechvibe.com/insights/staff-articles/how-ai-is-improving-predictive-analytics/

Sledge, P. (2020, September 23). *What Is The Social Dilemma: 6 Things To Know About The Netflix Documentary Before You Watch*. CINEMABLEND. https://www.cinemablend.com/news/2555346/what-is-the-social-dilemma-things-to-know-about-the-netflix-documentary-before-you-watch

Society of Actuaries. (2020). *Exam PA: Predictive Analytics | SOA*. SOA. https://www.soa.org/education/exam-req/edu-exam-pa-detail/

Soni, D. (2020, July 21). *Supervised vs. Unsupervised Learning - Towards Data Science*. Medium. https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d