Summative Assessment Report

| 1 | Introduction |
| --- | --- |
| 1.1 | Hypotheses and Research Question |
| 1.2 | Research Design |
| 1.3 | Methods |
| 2 | Recommendations For Paper Two |
| 2.1 | Use of Data Analytics |
| 2.2 | Philosophical Considerations |
| 2.3 | Ethical Considerations |
| 2.4 | Outline of the Research Design |
| 2.5 | Philosophy |
| 2.6 | Design Choice |
| 2.7 | Time Horizon |
| 2.8 | Technique and Procedure |
| 3 | Characteristics of Qualitative and Quantitative Methodologies |
| 4 | Conclusion |

Word count - 2948

1 Introduction

Paper 1[[1]](https://www.zotero.org/google-docs/?7WbTAH) and Paper 2 [[2]](https://www.zotero.org/google-docs/?DFZkFm) will be analysed to consider the research methods and methodologies used in their creation. The hypothesis and research will be identified and how the methods align to each methodology, and research outcome, will be analysed. Next, the research outcome of Paper 2 [[2]](https://www.zotero.org/google-docs/?hCyCYb) will be considered and the research question refined, alongside discussion of methods and methodology, to meet a similar research outcome. Lastly, quantitative and qualitative methodologies will be discussed.

1.1 Hypothesis and Research Questions

Paper 1’s research question is to predict energy use in a low energy house and how effective the proposed computational artificial intelligence model is [[1]](https://www.zotero.org/google-docs/?D8HwmO). A sub research question can be identified as which algorithm will predict the most accurate result. The dependent variable is identified as energy use for lights, while independent variables are temperature, wind speed and humidity. This is an applied research question as knowledge is used to solve a practical issue in computer science [[3]](https://www.zotero.org/google-docs/?EYPwS0). The hypothesis is that the proposed model, based on Support Vector Regression (SVR) and Artificial Neural Network with back propagation (BP-ANN), is able to predict electric use in a low energy house effectively [[1]](https://www.zotero.org/google-docs/?57Dx4o).

Paper 2’s hypothesis is can Big Data Analytics (BDA) be applied to criminal data, the inferred independent variable, in order to detect patterns and trends [[2]](https://www.zotero.org/google-docs/?Wi9iCg). The research questions are identified as: How can BDA be used in assisting police departments in understanding crime incidents, preventing incidents and optimising decision making. A further research question, which narrows the focus of the paper, is which machine learning and deep learning techniques, inferred dependent variables, are best for analysis and visualisation of criminal data [[2]](https://www.zotero.org/google-docs/?o9pYa2). Importantly, by narrowing the focus of the research question this paper builds on, and expands, existing knowledge in Computer Science. Similar to Paper 1 these research questions are applied research. Moreover, Paper 2’s research questions also follow an experimental design in order to discover and test the hypothesis [[4]](https://www.zotero.org/google-docs/?dmnyE2). Both papers use a quantitative methodology as the relationship between variables are investigated.

1.2 Research Design

The philosophy for both Paper 1 and Paper 2 align to a scientific approach rooted in epistemology, also known as positivism, as numerical data is used to prove a hypothesis [[5]](https://www.zotero.org/google-docs/?2lB8bd). Saunders highlights that by developing new knowledge, assumptions are made which shape how a researcher understands their research questions and choice of methods [[6]](https://www.zotero.org/google-docs/?sTu2sd). With this in mind, both papers make observations and analyse data in order to prove their hypothesis. Furthermore, theories in both papers are developed through the use of deduction. For instance, research begins with an existing theory, developed through literature reviews, to which a new hypothesis is presented [[7]](https://www.zotero.org/google-docs/?KcEXBd). Data is then collected in order to prove or disprove each paper’s respective hypothesis. Notably, secondary data sources are used in both papers for the training, and testing, of their respective predictive models [1][2].

1.3 Methods

Paper 1 [[1]](https://www.zotero.org/google-docs/?i9Ie5T) has utilised SVR, a machine learning method, alongside an Artificial Neural Network with back propagation (BP-ANN). SVR can be applied to regression problems where the objective is to predict a continuous output variable. Indeed, SVR is robust for outliers and small data sets [[8]](https://www.zotero.org/google-docs/?vKRlrY) which makes it a suitable quantitative method for Paper 1’s research question and methodology. A disadvantage is SVR’s sensitivity to kernel selection and potential for overfitting [[9]](https://www.zotero.org/google-docs/?UvfFVA). Importantly, SVR supports the research aims as the kernel tricks were adapted and used to identify patterns to predict results with fewer errors [1]. Interestingly, BP-ANN was found to outperform SVR in Paper 1 [1] which is similar to another earlier quantitative study [[10]](https://www.zotero.org/google-docs/?mXncyB) and also concluded BP-ANN is slightly more effective than SVR. Therefore, an alternative method to SVR is a Random Forest (RF) which has the advantage of working well with an increased number of data features and has a reduced risk of over-fitting. Indeed, [[11]](https://www.zotero.org/google-docs/?iui3OO) argues that RFs outperform similar models. However, RFs can be time-consuming on large data sets and require more computational resources [[12]](https://www.zotero.org/google-docs/?pSU4EO).

Paper 2 uses a LSTM model which has benefits in sequence modelling, such as capturing long-term contextual information and the ability to fit non-linearity, but the modelling effect is general and computation is time-consuming [[13]](https://www.zotero.org/google-docs/?YKikuG). An alternative method is Gated Recurrent Unit (GRU) which has fewer parameters, so it is faster to train with improved network performance [[14]](https://www.zotero.org/google-docs/?sF4GxT). The Prophet model was also used which is advantageous with time series data, seasonal trends and when non-linear trends are fit weekly or yearly [[15]](https://www.zotero.org/google-docs/?54uu5c). Moreover, [15] argues that Prophet provides more accurate predictions than an alternative model, Auto-Regressive Integrated Moving Averages (ARIMA), when the parameters are automatic. However, ARIMA could be a viable alternative if seasonality does not need to be considered [[16]](https://www.zotero.org/google-docs/?S4v8Jo) [[17]](https://www.zotero.org/google-docs/?xBQhVt). Evidently [[18]](https://www.zotero.org/google-docs/?OQg19X) strongly suggests that ARIMA, with a LSTM model, outperformed Prophet due to a lack of seasonal data. Significantly, given the research aims and consideration of seasonality in [2] Prophet is a suitable method for the data and research questions in comparison to ARIMA.

2 Recommendations For Paper Two

A new starting point for the outcome of Paper 2 is a mixed methods methodology to use quantitative data and open-ended qualitative data to provide a similar outcome. The new research question is: How can BDA be used to assist police departments' understanding of crime and identify the social issues causing crime in order to understand and prevent future incidents.

2.1 Use of Data Analytics

A key area of interest in Paper 2 is BDA which aligns to the research questions and the quantitative methodology. BDA allows for finding patterns and trends [[19]](https://www.zotero.org/google-docs/?xeHpkF), but it does have areas which need consideration, such as data analysis, when using prediction analysis to prevent incidents.

For instance, predictive analytics are rooted in statistical methods. Notably, the use of statistical significance, a small sample representing the population, may not be applicable [[20]](https://www.zotero.org/google-docs/?XIqcoM). This is due to the size of Big Data samples which can be the entire population, such as Paper 2’s crime data [2]. In contrast to traditional statistical analysis, predictive analysis does not describe the data but makes predictions from it. Therefore, methods, such as deep and machine learning methods, have been designed to address disadvantages in traditional statistical analytics with Big Data. An advantage is these methods can be used in an architectural modular way [[21]](https://www.zotero.org/google-docs/?Z4QeTn) giving researchers flexibility to consider specific, or combined, methods. Indeed, Paper 2 [2] uses Prophet and LSTM which, given the research question, are appropriate for predictive analysis and do not need refinement for addressing incident prevention in the research outcome.

That being said, a similar study was conducted by Kumar et al and used a k-means clustering method [[22]](https://www.zotero.org/google-docs/?1TQb3g). Similarly, a quantitative methodology was used as well as publicly available data on crime. However, different methods were used to reach their research outcome of visualising crime. K-means was a suitable choice as data was segmented and exploratory data analysis was carried out. With this in mind, Prophet and LSTM methods strongly align to the refined research question as predictions from the data are required.

2.2 Philosophical Considerations

The underlying philosophy for Paper 2’s research question is a scientific approach rooted in positivism, which is appropriate for a quantitative methodology. For example, [[5]](https://www.zotero.org/google-docs/?LGq7KQ) argues that a positivism philosophy requires quantifiable and statistical analysis. However, BDA presents epistemological challenges for scientific philosophy as there is a change towards using data to reach data-driven discoveries, rather than starting with samples or theories [[23]](https://www.zotero.org/google-docs/?uG5E7Y). Notably, the Streetlight effect, where a phenomena is studied and the relevant problems are not investigated [[24]](https://www.zotero.org/google-docs/?4kKN8z), highlights the need for rigorous sampling and selection of data before drawing conclusions. With this in mind, a starting point for this research could be pragmatism, under a mixed methods design, which is arguably a highly suitable philosophy to combine positivism and interpretivism [[25]](https://www.zotero.org/google-docs/?Bpkzmf). Significantly, pragmatism assists acknowledging bias in both data and the researcher alongside combining quantitative and qualitative methods.

2.3 Ethical Considerations

Ethical implications of the research question and outcome should be considered when using BDA to create predictive models. Concerningly, data to train predictive crime models can reflect biassed or corrupt policing practices [[26]](https://www.zotero.org/google-docs/?vOnXXV) meaning the outputted results will perpetuate any inputted biassed data. For instance, a predictive crime model, PredPol, was removed from use in Los Angeles [[27]](https://www.zotero.org/google-docs/?0S84nW) due to ethical and racial group bias concerns. Supporting this argument Campedelli [[28]](https://www.zotero.org/google-docs/?LNyeUQ) discovered a lack of research on algorithmic discrimination and bias. On the other hand, [[29]](https://www.zotero.org/google-docs/?Pomyul) reports that they found no significant differences in ethinc-racial group arrests from a study in Los Angeles. Importantly, this demonstrates that careful selection and use of data is required to mitigate any potential bias. One approach to address this is inclusion of fewer crime categories when analysing big data [[30]](https://www.zotero.org/google-docs/?uvJ0G7). Arguably, this can also simplify analysis, enabling refinement of inputted data, potentially reducing bias. Therefore, a starting point for ethical consideration in the refined research outcome is: can specific crimes be predicted?

An argument for a qualitative approach to the research outcome is given by Neiva et al who investigated how police officers view and expect BDA to assist in criminal investigations [[31]](https://www.zotero.org/google-docs/?x8S5xo). Importantly, the research allows for a view from the persons directly affected by predictive models. Importantly, it gives insight to the impact of technology from the participant’s view and allows participants of the refined study to highlight if they feel policing discriminates as well as why a crime was committed.

2.4 Outline of the Research Design

Using a mixed methods design, quantitative and qualitative methodologies are combined in order to gain a broad and in depth research outcome. Notably, the qualitative methodology will enable a deeper understanding of why crime is being committed, and quantitative methods allow for the numerical collection and analysis of data [[32]](https://www.zotero.org/google-docs/?Updwgl). A mixed methods approach is suitable when a quantitative or qualitative methodology is insufficient to best comprehend a problem, or a detailed understanding is desired [[33]](https://www.zotero.org/google-docs/?65pgHd). Furthermore, one methodology's strengths can compensate for the other's weakness, allowing questions to be answered which may not necessarily be achievable by one methodology alone [[34]](https://www.zotero.org/google-docs/?SDkqdc). However, mixed methods can be time consuming and a researcher needs to be confident in quantitative and qualitative methodologies.

In creating the research design, Saunders’ Onion model [[6]](https://www.zotero.org/google-docs/?Wq8tlQ) has been utilised. Philosophy, approach, choice, time horizon and techniques and procedures will be examined.

2.5 Philosophy

Positivism and interpretivism, individuals experience object reality, align to quantitative and qualitative methodologies respectively. With mixed methods, pragmatism allows the researcher to focus on the research problem and use different approaches [[33]](https://www.zotero.org/google-docs/?bt6w13) which makes pragmatism suitable for a mixed methods research design. Significantly, Feilzer [[35]](https://www.zotero.org/google-docs/?hX5gsh) argues, through the example of crime data, pragmatism supports different research methods and analysis techniques. This view is also supported by [[36]](https://www.zotero.org/google-docs/?XdEwuf) who also state that inductive, deductive and abductive reasoning can aid in creating knowledge. Importantly, pragmatism supports the use of different methodologies and, ideologically, researcher and participant’s values are important to the research design with biases acknowledged through axiology [[6]](https://www.zotero.org/google-docs/?mExQoR).

2.6 Design Choice

There are two main design choices for mixed methods, sequential and concurrent. For this research a sequential design, quantitative then qualitative, has been chosen as it allows for quantitative data collection first and then qualitative data to elaborate the results [[37]](https://www.zotero.org/google-docs/?9Ei0Vm). Importantly, inference can be drawn from the integration of both results. Advantages for a sequential method are that it is straightforward to implement and there is an opportunity for the investigation of quantitative results in detail [[38]](https://www.zotero.org/google-docs/?YgyErK). However, a limitation is the breadth of time for the study and re-establishing contact with participants if required. Unlike a qualitative then quantitative sequential design this approach allows for the researcher to explore the data, and methods, for prediction modelling first.

2.7 Time Horizon

Time horizon refers to the period of time the research will take. As the data is not being collected over an extended time period [[34]](https://www.zotero.org/google-docs/?77guFN) the time horizon for this research is cross-sectional.

2.8 Technique and Procedure

Quantitative and qualitative techniques for data collection and analysis will be used. The quantitative techniques will remain unchanged for collection of public data [2]. However, one consideration is missing data or outliers [[39]](https://www.zotero.org/google-docs/?2xi6rZ), however the Prophet model is well suited to handle these cases.

In order to address the refined research question, a focus will be placed on looking at specific crime categories. Recent research suggests dividing crime categories into separate groups so analysis is easier, instead of focusing on all crime at once [[30]](https://www.zotero.org/google-docs/?ZXCdRR). For instance, physical assault or robbery could be selected and then be followed up with qualitative interviews with arrested persons and police officers.

The qualitative techniques will be interviews, which will be informal and open-ended so as to gain an in-depth understanding [[40]](https://www.zotero.org/google-docs/?cbGwKZ). This will allow for an insight into why an individual committed a crime alongside a view from police officers. Importantly, interviews will try to uncover if crimes are premeditated, spontaneous or committed out of economical circumstances. Interviews can then be analysed against the quantitative data to detect any patterns, such as a factory closing down, and an increase in crime in that area from economic hardship. This could then be a factor in the quantitative prediction model. A final consideration is anonymity and confidentiality of participants.

3 Characteristics of Qualitative and Quantitative Methodologies

Quantitative research designs consider measurable data, such as dependent and independent variables, from experiments or surveys and focus on objectivity. On the other hand, qualitative research designs explore meaning and interpretations, such as an interview about an event, or other open-ended methods in order to discover in depth information to understand a problem [[32]](https://www.zotero.org/google-docs/?QJzpQw).

A key strength of a quantitative methodology is standardised instruments are used for data collection, meaning replication is possible and large samples can be used [[41]](https://www.zotero.org/google-docs/?ffkEnd). However, weakness can include structural bias, missing data or inappropriate sampling, and a lack of context [[41]](https://www.zotero.org/google-docs/?Bcy3Es). Whereas, qualitative methodologies strengths are meaningful insights, such as a detailed description of experience, and the generation of new ideas through open-ended responses. However, weakness can include research biases, such as the Hawthorne effect and recall bias, as well as being labour intensive [[33]](https://www.zotero.org/google-docs/?oZfPN4).

An example of a quantitative is experimentation where an independent variable is controlled or manipulated to measure effects on a dependent variable. An advantage of this method is the environment, where research takes place, can be controlled so it is easier to estimate the effects of the variable [[42]](https://www.zotero.org/google-docs/?JIDwrn). Contrastingly, a disadvantage is that it can be difficult to externally validate the experiment due to the artificial nature of the experiment’s context. For example, [[43]](https://www.zotero.org/google-docs/?eBDsG6) experiments are conducted to achieve higher efficiency and accurate results when compared to state-of-the-art convolutional neural networks. Importantly, evaluations of the experimental design were tested against real-world databases of images to strengthen the results of their experiments. This shows considerations to the weakness of an experimental method by rigorous testing with uncontrolled variables. Indeed, this method’s strength is testing a hypothesis.

An alternative method is a survey, which asks questions from a group of participants [[44]](https://www.zotero.org/google-docs/?i0Y2qw). An advantage is surveys are a practical method for collecting data and are scalable [[45]](https://www.zotero.org/google-docs/?vLMMn6). However, participants may give inaccurate answers and surveys do not capture emotion from participants. Arguably, surveys are a fast way of data collection, and can be used for this, but can be inaccurate if not engaged well with by participants.

An example of a qualitative method are case studies which can generate detailed information about individuals as well as offering the potential for innovation [[33]](https://www.zotero.org/google-docs/?44Pzat). A disadvantage is that establishing cause and effect connections can be difficult alongside ethical issues which may arise. For example, [[46]](https://www.zotero.org/google-docs/?JbrhwL) used case studies to draw attention to the issues of artificial learning and data analytics and highlight where existing ethical frameworks fail. A strength of this approach was identifying areas of concern and using innovation to create a pilot database. However, a weakness is generalisation and a single connection, even in one case study, may be theoretically important [[47]](https://www.zotero.org/google-docs/?cPNNMI). Importantly, a small sample size was used to be effective and maximise the success of the case study method.

Alternatively, a focus group is another qualitative method for data collection where participants are brought together in a small group to answer questions [[48]](https://www.zotero.org/google-docs/?WiIejR). This method has the advantages of being inexpensive and results being more comprehensive than raw data. However, participants have less time to speak than in an interview, and some participants may dominate the conversation. Therefore, focus groups should be used when a diverse set of responses are required for data collection.

4 Conclusion

In conclusion, good research has a narrow focus on a specific problem, or area of interest, to which new knowledge is discovered. Two methodologies, quantitative and qualitative, enable different approaches for the types, and methods, for data collection and analysis. Mixed methods presents a hybrid methodology for when an area of research needs a quantitative and qualitative approach. Importantly, mixed methods allow for both depth and explanation of a problem or phenomenon. Specifically, Computer Science can benefit from a mixed methods approach as it is based in mathematics and science but can have a significant impact on people.

Different methods have advantages and disadvantages so it is important for the researcher to carefully consider which will be best for their research outcomes. Moreover, a researcher's personal philosophy and worldview can influence their research and chosen approaches. It is therefore important that planning research all these aspects, delineated in Saunders’s Onion, are considered. This will result in focused and well supported research to add to the knowledge of humanity.

Reference List

[[1] R. I. Rasel, N. Sultana, S. Akther, and A. Haroon, ‘Predicting Electric Energy Use of a Low Energy House: A Machine Learning Approach’, in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Cox’sBazar, Bangladesh: IEEE, Feb. 2019, pp. 1–6. doi: 10.1109/ECACE.2019.8679479.](https://www.zotero.org/google-docs/?WY6Huu)

[[2] M. Feng *et al.*, ‘Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data’, *IEEE Access*, vol. 7, pp. 106111–106123, 2019, doi: 10.1109/ACCESS.2019.2930410.](https://www.zotero.org/google-docs/?WY6Huu)

[[3] M. D. Gall and W. R. Borg, *Applying educational research*, 7th ed. Hoboken: Pearson, 2015.](https://www.zotero.org/google-docs/?WY6Huu)

[[4] S. L. T. McGregor, *Understanding and Evaluating Research: A Critical Guide*. SAGE Publications, Inc, 2018. doi: 10.4135/9781071802656.](https://www.zotero.org/google-docs/?WY6Huu)

[[5] M. M. Al-Ababneh, ‘Linking Ontology, Epistemology And Research Methodology’, *Sci. Philos.*, vol. 8, no. 1, Jul. 2020, doi: 10.23756/sp.v8i1.500.](https://www.zotero.org/google-docs/?WY6Huu)

[[6] M. N. K. Saunders, P. Lewis, and A. Thornhill, *Research methods for business students*, Eighth Edition. New York: Pearson, 2019.](https://www.zotero.org/google-docs/?WY6Huu)

[[7] N. Blaikie, *Designing Social Research*, 2nd ed. Cambridge: Polity, 2010.](https://www.zotero.org/google-docs/?WY6Huu)

[[8] F. Zhang and L. O’Donnell, ‘Chapter 7 Support Vector Regression’, in *Methods and Applications to Brain Disorders*, Academic Press, 2020, pp. 123–140. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128157398000079](https://www.zotero.org/google-docs/?WY6Huu)

[[9] A. J. Smola and B. Schölkopf, ‘A tutorial on support vector regression’, *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004, doi: 10.1023/B:STCO.0000035301.49549.88.](https://www.zotero.org/google-docs/?WY6Huu)

[[10] S. Nurunnahar, D. B. Talukdar, R. I. Rasel, and N. Sultana, ‘A short term wind speed forcasting using SVR and BP-ANN: A comparative analysis’, in *2017 20th International Conference of Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh: IEEE, Dec. 2017, pp. 1–6. doi: 10.1109/ICCITECHN.2017.8281802.](https://www.zotero.org/google-docs/?WY6Huu)

[[11] M. Fernandez-Delgado, E. Cernadas, S. Barro, and D. Amorim, ‘Do we Need Hundreds of Classiﬁers to Solve Real World Classiﬁcation Problems?’.](https://www.zotero.org/google-docs/?WY6Huu)

[[12] ‘What is random forest?’, Apr. 21, 2023. https://www.ibm.com/topics/random-forest](https://www.zotero.org/google-docs/?WY6Huu)

[[13] G. Memarzadeh and F. Keynia, ‘A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets’, *Energy Convers. Manag.*, vol. 213, p. 112824, Jun. 2020, doi: 10.1016/j.enconman.2020.112824.](https://www.zotero.org/google-docs/?WY6Huu)

[[14] K. E. ArunKumar, D. V. Kalaga, Ch. Mohan Sai Kumar, M. Kawaji, and T. M. Brenza, ‘Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends’, *Alex. Eng. J.*, vol. 61, no. 10, pp. 7585–7603, Oct. 2022, doi: 10.1016/j.aej.2022.01.011.](https://www.zotero.org/google-docs/?WY6Huu)

[[15] S. J. Taylor and B. Letham, ‘Forecasting at scale’, PeerJ Preprints, preprint, Sep. 2017. doi: 10.7287/peerj.preprints.3190v2.](https://www.zotero.org/google-docs/?WY6Huu)

[[16] D. Duarte and J. Faerman, ‘Comparison of Time Series Prediction of Healthcare Emergency Department Indicators with ARIMA and Prophet’, in *9th International Conference on Computer Science, Engineering and Applications (ICCSEA 2019)*, Aircc publishing Corporation, Dec. 2019, pp. 123–133. doi: 10.5121/csit.2019.91810.](https://www.zotero.org/google-docs/?WY6Huu)

[[17] L. Menculini *et al.*, ‘Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices’, *Forecasting*, vol. 3, no. 3, pp. 644–662, Sep. 2021, doi: 10.3390/forecast3030040.](https://www.zotero.org/google-docs/?WY6Huu)

[[18] Y. Ning, H. Kazemi, and P. Tahmasebi, ‘A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet’, *Comput. Geosci.*, vol. 164, p. 105126, Jul. 2022, doi: 10.1016/j.cageo.2022.105126.](https://www.zotero.org/google-docs/?WY6Huu)

[[19] P. Mikalef, I. O. Pappas, J. Krogstie, and M. Giannakos, ‘Big data analytics capabilities: a systematic literature review and research agenda’, *Inf. Syst. E-Bus. Manag.*, vol. 16, no. 3, pp. 547–578, Aug. 2018, doi: 10.1007/s10257-017-0362-y.](https://www.zotero.org/google-docs/?WY6Huu)

[[20] R. H. Hariri, E. M. Fredericks, and K. M. Bowers, ‘Uncertainty in big data analytics: survey, opportunities, and challenges’, *J. Big Data*, vol. 6, no. 1, p. 44, Dec. 2019, doi: 10.1186/s40537-019-0206-3.](https://www.zotero.org/google-docs/?WY6Huu)

[[21] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, ‘An Introductory Review of Deep Learning for Prediction Models With Big Data’, *Front. Artif. Intell.*, vol. 3, p. 4, Feb. 2020, doi: 10.3389/frai.2020.00004.](https://www.zotero.org/google-docs/?WY6Huu)

[[22] A. V. Kumar, S. Chitumadugula, and V. T. Rayalacheruvu, ‘Crime Data Analysis using Big Data Analytics and Visualization using Tableau’, in *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, Coimbatore, India: IEEE, Dec. 2022, pp. 627–632. doi: 10.1109/ICECA55336.2022.10009119.](https://www.zotero.org/google-docs/?WY6Huu)

[[23] A. Elragal and R. Klischewski, ‘Theory-driven or process-driven prediction? Epistemological challenges of big data analytics’, *J. Big Data*, vol. 4, no. 1, p. 19, Dec. 2017, doi: 10.1186/s40537-017-0079-2.](https://www.zotero.org/google-docs/?WY6Huu)

[[24] A. Elragal and M. Haddara, ‘Design Science Research: Evaluation in the Lens of Big Data Analytics’, *Systems*, vol. 7, no. 2, p. 27, May 2019, doi: 10.3390/systems7020027.](https://www.zotero.org/google-docs/?WY6Huu)

[[25] H. Maarouf, ‘Pragmatism as a Supportive Paradigm for the Mixed Research Approach: Conceptualizing the Ontological, Epistemological, and Axiological Stances of Pragmatism’, *Int. Bus. Res.*, vol. 12, no. 9, p. 1, Aug. 2019, doi: 10.5539/ibr.v12n9p1.](https://www.zotero.org/google-docs/?WY6Huu)

[[26] R. Richardson, J. M. Schultz, and K. Crawford, ‘DIRTY DATA, BAD PREDICTIONS: HOW CIVIL RIGHTS VIOLATIONS IMPACT POLICE DATA, PREDICTIVE POLICING SYSTEMS, AND JUSTICE’, *N. Y. Univ. LAW Rev.*, vol. 94.](https://www.zotero.org/google-docs/?WY6Huu)

[[27] J. Bhuiyan, ‘LAPD ended predictive policing programs amid public outcry. A new effort shares many of their flaws’, *The Guardian*, Nov. 08, 2021.](https://www.zotero.org/google-docs/?WY6Huu)

[[28] G. M. Campedelli, ‘Where Are We? Using Scopus to Map the Literature at the Intersection Between Artificial Intelligence and Research on Crime’, *J. Comput. Soc. Sci.*, vol. 4, no. 2, pp. 503–530, Nov. 2021, doi: 10.1007/s42001-020-00082-9.](https://www.zotero.org/google-docs/?WY6Huu)

[[29] P. J. Brantingham, M. Valasik, and G. O. Mohler, ‘Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial’, *Stat. Public Policy*, vol. 5, no. 1, pp. 1–6, Jan. 2018, doi: 10.1080/2330443X.2018.1438940.](https://www.zotero.org/google-docs/?WY6Huu)

[[30] M. Mansour Salah and K. Xia, ‘Big Crime Data Analytics and Visualization’, in *2022 The 6th International Conference on Compute and Data Analysis*, Shanghai China: ACM, Feb. 2022, pp. 24–28. doi: 10.1145/3523089.3523094.](https://www.zotero.org/google-docs/?WY6Huu)

[[31] L. Neiva, R. Granja, and H. Machado, ‘Big Data applied to criminal investigations: expectations of professionals of police cooperation in the European Union’, *Polic. Soc.*, vol. 32, no. 10, pp. 1167–1179, Nov. 2022, doi: 10.1080/10439463.2022.2029433.](https://www.zotero.org/google-docs/?WY6Huu)

[[32] A. Queirós, D. Faria, and F. Almeida, ‘Strengths And Limitations Of Qualitative And Quantitative Research Methods’, Sep. 2017, doi: 10.5281/ZENODO.887089.](https://www.zotero.org/google-docs/?WY6Huu)

[[33] J. Creswell and J. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*, 5th ed. Thousand Oaks, CA: SAGE Publications, 2018.](https://www.zotero.org/google-docs/?WY6Huu)

[[34] J. Creswell and V. L. Plano, *Designing and Conducting Mixed Methods Research*, Third Edition. California: SAGE Publications, 2018.](https://www.zotero.org/google-docs/?WY6Huu)

[[35] M. Yvonne Feilzer, ‘Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm’, *J. Mix. Methods Res.*, vol. 4, no. 1, pp. 6–16, Jan. 2010, doi: 10.1177/1558689809349691.](https://www.zotero.org/google-docs/?WY6Huu)

[[36] A. Mitchell, ‘A Review of Mixed Methods, Pragmatism and Abduction Techniques’, vol. 16, no. 3, 2018.](https://www.zotero.org/google-docs/?WY6Huu)

[[37] V. L. Plano Clark and N. V. Ivankova, ‘Chapter 5: How to use Mixed Methods Research?: Understanding the Basic Mixed Methods Designs’, in *Mixed Methods Research: A Guide to the Field*, 2455 Teller Road, Thousand Oaks California 91320: SAGE Publications, Inc., 2016. doi: 10.4135/9781483398341.](https://www.zotero.org/google-docs/?WY6Huu)

[[38] D. Morgan, *Integrating Qualitative and Quantitative Methods: A Pragmatic Approach.* SAGE Publications, 2013.](https://www.zotero.org/google-docs/?WY6Huu)

[[39] N. Macià and E. Bernadó-Mansilla, ‘Towards UCI+: A mindful repository design’, *Inf. Sci.*, vol. 261, pp. 237–262, Mar. 2014, doi: 10.1016/j.ins.2013.08.059.](https://www.zotero.org/google-docs/?WY6Huu)

[[40] D. Neuman, ‘Qualitative research in educational communications and technology: a brief introduction to principles and procedures’, *J. Comput. High. Educ.*, vol. 26, no. 1, pp. 69–86, Apr. 2014, doi: 10.1007/s12528-014-9078-x.](https://www.zotero.org/google-docs/?WY6Huu)

[[41] Faculty of Arts and Social Science, University of Malaya, Malaysia and L. T. Choy, ‘The Strengths and Weaknesses of Research Methodology: Comparison and Complimentary between Qualitative and Quantitative Approaches’, *IOSR J. Humanit. Soc. Sci.*, vol. 19, no. 4, pp. 99–104, 2014, doi: 10.9790/0837-194399104.](https://www.zotero.org/google-docs/?WY6Huu)

[[42] A. F. Hayes and K. J. Preacher, ‘Statistical mediation analysis with a multicategorical independent variable’, *Br. J. Math. Stat. Psychol.*, vol. 67, no. 3, pp. 451–470, Nov. 2014, doi: 10.1111/bmsp.12028.](https://www.zotero.org/google-docs/?WY6Huu)

[[43] Y. Wang, B. Feng, X. Peng, and Y. Ding, ‘An Efficient Quantitative Approach for Optimizing Convolutional Neural Networks’, in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event Queensland Australia: ACM, Oct. 2021, pp. 2050–2059. doi: 10.1145/3459637.3482230.](https://www.zotero.org/google-docs/?WY6Huu)

[[44] S. Sukamolson, ‘Fundamentals of Quantitative Research’.](https://www.zotero.org/google-docs/?WY6Huu)

[[45] M. S. Rahman, ‘The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language “Testing and Assessment” Research: A Literature Review’, *J. Educ. Learn.*, vol. 6, no. 1, p. 102, Nov. 2016, doi: 10.5539/jel.v6n1p102.](https://www.zotero.org/google-docs/?WY6Huu)

[[46] K. Kitto and S. Knight, ‘Practical ethics for building learning analytics’, *Br. J. Educ. Technol.*, vol. 50, no. 6, pp. 2855–2870, Nov. 2019, doi: 10.1111/bjet.12868.](https://www.zotero.org/google-docs/?WY6Huu)

[[47] A. B. Starman, ‘The case study as a type of qualitative research’.](https://www.zotero.org/google-docs/?WY6Huu)

[[48] M. Hancock, L. Amankwaa, M. Revell, and D. Mueller, ‘Focus Group Data Saturation: A New Approach to Data Analysis’, *Qual. Rep.*, Nov. 2016, doi:10.46743/2160-3715/2016.2330.](https://www.zotero.org/google-docs/?WY6Huu)