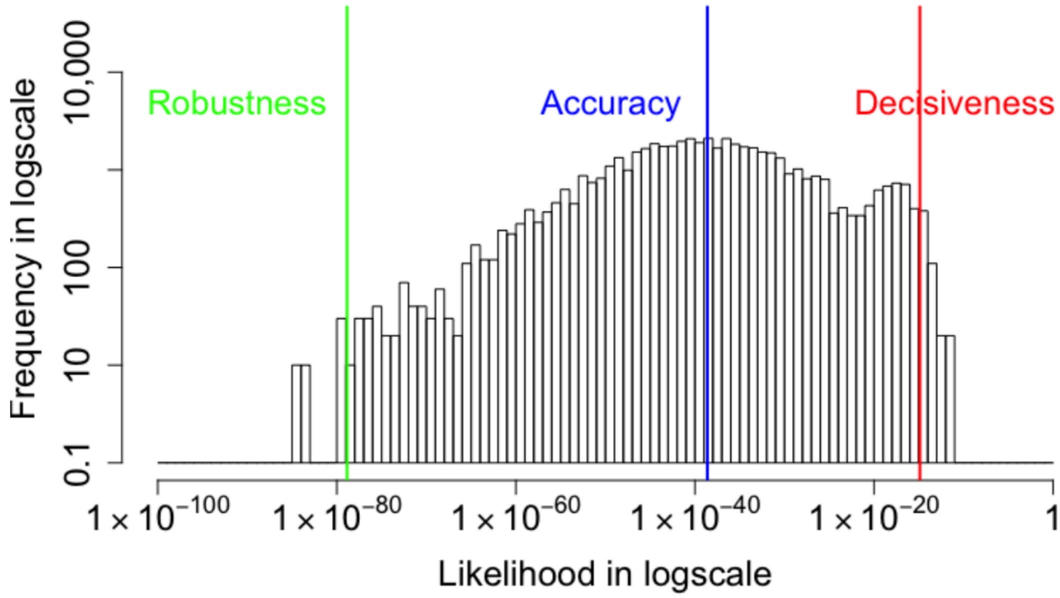
SingularityNET Risk-Aware Projects

Deep Fund 1: Report 4, Risk Assessment Final Report

Assessment of and Data Generation for AI Applications



PI: Kenric P. Nelson, Kenric.Nelson@photrek.io
Co-PI: William Thistleton
Contributors: Hong Yue, Ethan Nelson, Blake Anderton

02.03.2023

PHOTREK

We Enlighten Your Path

EXECUTIVE SUMMARY

This is the final report for Photrek's delivery of the Risk-Aware Assessment Service for SingularityNET's AI Marketplace. This is R1-B2-Milestone 4 (Final Report, $5,500) R1-B2-Milestone 5 (Hosting Costs, $7,500). In the introduction we provide a breakdown of the funds allocated for the project and how they were assigned to each milestone. Next we explain the algorithm, its purpose and its theoretical basis in information theory. We also demonstrate applications of this algorithm to another Deep Fund project, namely the Risk-Aware Data Generator, as well as in assessing FiveThirtyEight's 2022 midterm elections. The final section gives a breakdown of the technical accomplishments for each milestone as well as the obstacles we faced along the way.

INTRODUCTION

On the 1st of May 2022, Photrek submitted a proposal to the SingularityNet Deep Fund to build an assessment service that would evaluate the performance of probabilistic machine learning models. Our service provides a histogram plot of the probabilities generated by the model and overlays the calculation of three metrics (Accuracy, Decisiveness and Robustness) which summarise different aspects of the prediction distribution (Nelson, 2017).

DESCRIPTION OF RISK ASSESSMENT SERVICE

Photrek's Risk Assessment Service draws upon two innovations in the application of information theory to the assessment of AI/ML algorithms, hereafter referred to as machine intelligence (MI) algorithms. First, while the mathematics of information theory relies on the logarithm of probabilities to form an additive space in which the arithmetic mean can be applied, we translate those results back to the probability space so that an average probability, we call "Accuracy" can be computed. The probability space is multiplicative and as such, the average probability is the geometric mean, i.e. the probabilities are multiplied and then a nth root is taken, where n is the number of probabilities.

Secondly, Photrek utilizes a generalization of information theory that models a degree of risk tolerance. Although a variety of generalizations to the entropy function have been proposed, such as the Renyi and Tsallis entropies, these and many others when translated to the probability domain form the generalized mean of the probabilities.

This report will not detail the mathematics but the assessment methodology was introduced in (Nelson, 2017), the tutorial (Nelson, 2021) provides a simplified overview, and (Nelson, 2017b) provides mathematical proofs.

Using the generalized mean (GM) of probabilities forecasted by a MI algorithm, Photrek computes a Robustness, Accuracy, and Decisiveness metric. The metrics are applied to a set of probabilities assigned to events that actually occurred during a test evaluation. As described earlier, the Accuracy metric is consistent with the log score of information theory and is computed by the geometric mean of the probabilities. The geometric mean is the zeroth power of the GM. The Decisiveness is measured by the arithmetic mean, which is the first power of the GM. Decisiveness is closely related to the classification performance of an algorithm; i.e. the performance of making decisions rather than the accuracy of the probabilities. Because Decisiveness is a weak measurement, it is necessary to include a stronger metric to make sure that an algorithm will be robust against deviations from the original training set distribution. To measure the Robustness, Photrek utilizes the -⅔ power of the GM. The negative power inverts the probabilities, thus giving more weight to low probabilities. This provides a metric of the performance of poor forecasts with low probability. The fraction two-thirds turns out to be the natural reciprocal of the arithmetic mean due to the properties of the generalized entropy, which is explained in the references.

TWO USE CASES OF THE RISK-AWARE ASSESSMENT

Photrek completed two demonstrations of the Risk-Aware Assessment. The assessment of image processing is reviewed here. The assessment of the forecast for the 2022 House Election was published on LinkedIn and is included as an appendix.

Data generation algorithms such as the Variational Autoencoder (VAE) produce complex data samples by decoding a learned statistical model. During the training and testing of these algorithms a probability likelihood of the generated image matching a training or test image is computed. In the scientific literature these results are commonly reported as the average (arithmetic) of log-likelihood of the reconstruction. Using Photrek's Risk Assessment methodology, we can provide the full histogram of the performance overlaid with the Decisiveness, Accuracy, and Robustness metrics.

Figures 1 and 2 are from Report 4 which included results on our other SingularityNET Deep Fund project, "Risk-Aware Data Generator for SingularityNET applications''. Figure

1 shows performance measurements of the Coupled VAE algorithm that incorporates into the training process the coupled entropy allowing for tunable risk tolerance. The MNIST handwritten numeral dataset was used to evaluate the algorithm. Using a coupling value of 0.5, we were able to generate images with a tight distribution. The accuracy is approximately 10^{-15}, which is significantly higher than the non-coupling VAE algorithms performance of 10^{-40} (see Report 4). Furthermore the Robustness is approximately 10^{-30}, which is dramatically less than the VAE robustness which was less than 10^{-90}.

Figure 2 evaluates the coupled VAE algorithm when it is seeded with an image corrupted with shot noise. There is only a small degradation in the performance and continues to produce images with higher likelihoods than the original VAE seeded with uncorrupted images. The Accuracy and Robustness are approximately 10^{-20} and 10^{-40}, respectively.

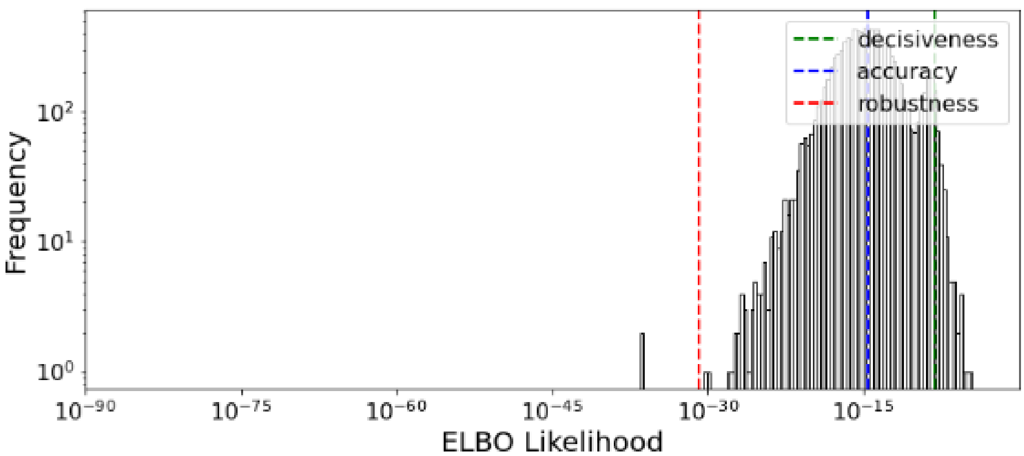


Figure 1 High Coupling .5, Latent Space Dim=32, Uncorrupted

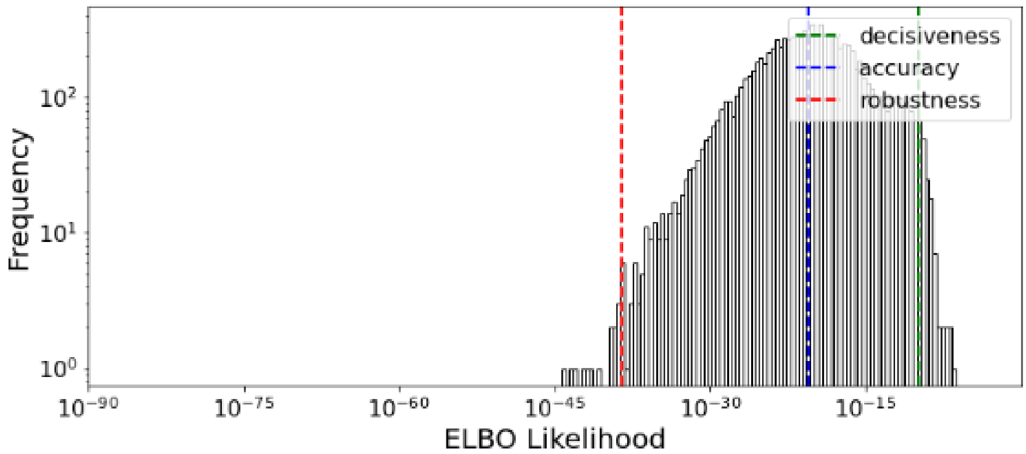


Figure 2 High Coupling .5, Latent Space Dim=32, Shot Noise

INTEGRATING DECENTRALISED APPLICATION TO THE SINGULARITYNET MARKETPLACE

Status

Photrek successfully progressed through the process of establishing an operational SNET service. These included (1) configuring and testing an appropriate cloud platform for hosting the service, daemon, and ETCD cluster; and (2) designing/testing an intuitive and explanatory User Interface (as a Marketplace-deployed Decentralized Application – DApp – see Figure 4) to engage and educate service users.

Explored Sustainable Configurations

In the process of integrating the service, Photrek obtained valuable information regarding hosting configurations for SNET services to reduce costs to service providers. These include cost-effectively configuring the service to facilitate “serverless” on-demand cloud processing (rather than execution on persistently available servers). Photrek anticipates these findings will increase sustainability for SNET service providers who rely upon cloud-hosted deployments.

Problems Viewed as Early-Adopter Progress

Photrek and the SingularityNet development team view this service’s deployment as an early-adopter exercise in progressing through SingularityNet’s existing registration/deployment work-flow. Our collaboration involved working through package-prerequisites not presently documented; developing verification criteria for component operation and validity of auxiliary files (such as SSL certificates); and determining alternative testing strategies due to the unavailability of active test-nets. Photrek's identification and communication of these findings supports streamlining new services' onboarding process, facilitating increased platform adoption and utilisation.

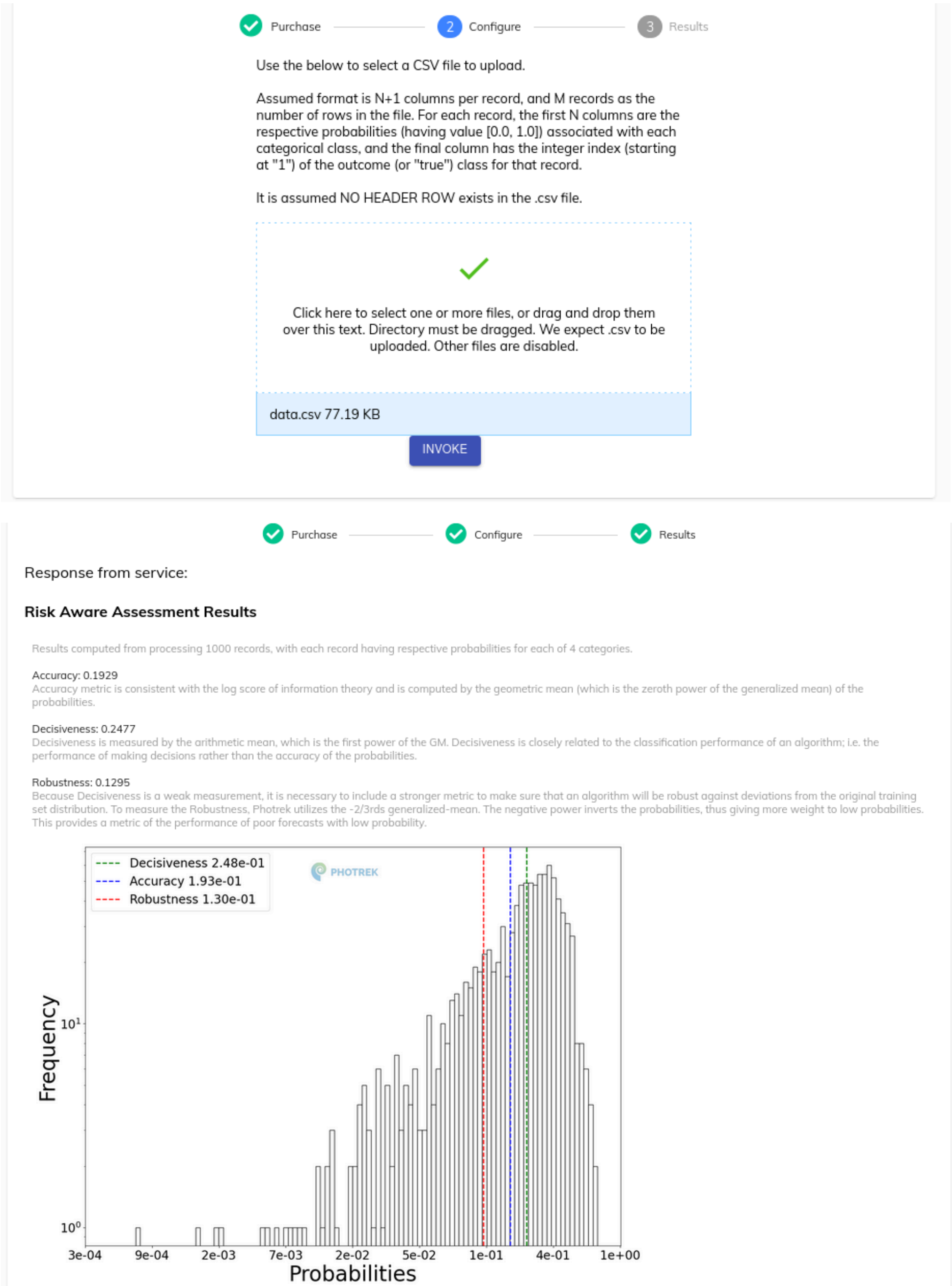


Figure 4: Demonstration of the Risk Aware Assessment DApp interface and execution results. Top illustrates the input interface; bottom illustrates the post-execution output delivered to end-users.

REVIEW OF MILESTONES

MILESTONE 1

For this milestone, Photrek was tasked with sketching out a design architecture for the Risk Assessment dApp. Photrek originally planned to use an Amazon Web Services platform to host the application and etcd key storage. However, during the second milestone the plan shifted to utilisation of Google Cloud, which would allow the service to integrate with the Workstation and Cloud computing service Photrek was already using for its business and machine learning requirements.

MILESTONE 2

Objective: provide a functioning web application that can connect to and operate on the SNET block chain.

Delivered: Photrek built a streamlit application that allowed users to upload a csv file of probabilities. The app would then display the histogram as well as calculate the ADR metrics. The service was hosted as a container on Google Kubernetes Engine (GKE). We also provided a GitHub repository that held our code for the calculations and the hosting of the server: https://github.com/Photrek/Risk-Aware-Assessment

We encountered technical difficulties at this stage which delayed connecting our service with the SNET marketplace.

MILESTONE 3

Objective: evaluate the performance of the assessment service.

Delivered: We made changes to the back end architecture to run using Google Cloud Functions instead of GKE. Cloud Functions gave us the advantage of only calling cloud compute when needed and not needing a server to be run at all times. We also built a separate front end that allowed us to transition away from a monolithic application design. This would allow future development and scaling to proceed faster.

Unfortunately, we still had issues with connecting to the SNET blockchain as we required Google SSL certificates to process transactions using the SNET Daemon. We initially hosted the Photrek website using GoDaddy but this did not allow us to properly connect with our Google Cloud back end, to do so required us to switch to a Google domain. This turned out to be a lengthy process and we were not able to make that switch before the report for the milestone was submitted.

MILESTONE 4

Objective: write a final report to summarise our progress and market the service to customers.

Delivered: The ADR metrics were used by Photrek to assess FiveThirtyEight's midterm election forecasts. We submitted a report explaining the workings of the algorithm and how FiveThirtyEight's forecasts performed in relation to the actual results. This report was published as a LinkedIn article and was included in a presentation to the 2022 Cardano Summit in Washington, DC.

This algorithm was also used internally as part of our Data Generator project to assess the quality of our reconstructions. For examples of this, please refer to our fourth report for deep fund 1.

Blake Anderton (Ph.D.) was brought on to assist in the integration of the service on the SNET platform, but the process proved to be much more difficult than anticipated. With many hours of development work and gratefully-acknowledged assistance from the SNET engineering team, we were able to successfully deploy an operational service (as illustrated in Figure 4).

BIBLIOGRAPHY

Cao, S., Li, J., Nelson, K. P., & Kon, M. A. (2022). Coupled VAE: Improved Accuracy and Robustness of a Variational Autoencoder. Entropy, 24(3). https://doi.org/10.3390/e24030423

George, C. A., Barrera, E. A., & Nelson, K. P. (2020). Applying the Decisiveness and Robustness Metrics to Convolutional Neural Networks. http://arxiv.org/abs/2006.00058

Nelson, K. P. (2017). Assessing Probabilistic Inference by Comparing the Generalized Mean of the Model and Source Probabilities. Entropy, 19(6), 286. https://doi.org/10.3390/e19060286

Nelson, K. P., Umarov, S. R., & Kon, M. A. (2017). On the average uncertainty for systems with nonlinear coupling. *Physica A: Statistical Mechanics and Its Applications*, *468*, 30–43. https://doi.org/10.1016/j.physa.2016.09.046

Nelson, K. P. (2021). Reduced Perplexity: A simplified perspective on assessing probabilistic forecasts. In M. Chen, J. M. Dunn, A. Golan, & A. Ullah (Eds.), Advances in Info-Metrics: Information and Information Processing across Disciplines. Oxford University Press. http://arxiv.org/abs/1603.08830

Nelson, K. P., Scannell, B. J., & Landau, H. (2011). A risk profile for information fusion algorithms. Entropy, 13(8), 1518–1532. https://doi.org/10.3390/e13081518

APPENDIX: Accuracy of FiveThirtyEight's 2022 House Election Forecasts

Originally published as a Linked-In article

Election forecasters get heavy criticized when the outcomes of elections deviate from expectations. Unfortunately, even when forecasts are accurate about the uncertainty of an election, public commentators may express more confidence in a particular outcome than is justified by the actual forecast. The 2022 US House election appears to be a case in point. In the week prior to the Tuesday, November 9th election, many news outlets were declaring the imminence of a "Red Wave" as predictions were made of large Republican wins. But did the forecasts of organizations like FiveThirtyEight really justify such an expectation?

Photrek's analysis of the ABC News FiveThirtyEight forecast for the US House of Representatives elections suggests that their probabilistic predictions were very accurate. This analysis is part of a project sponsored by SingularityNET to provide a service on their AI Marketplace to analyze the quality of machine learning forecasts. The accuracy of the probabilistic forecasts is more subtle to compute than the commonly referred to accuracy of the predicted outcome. And this is where confusion arises regarding interpretation of a forecast.

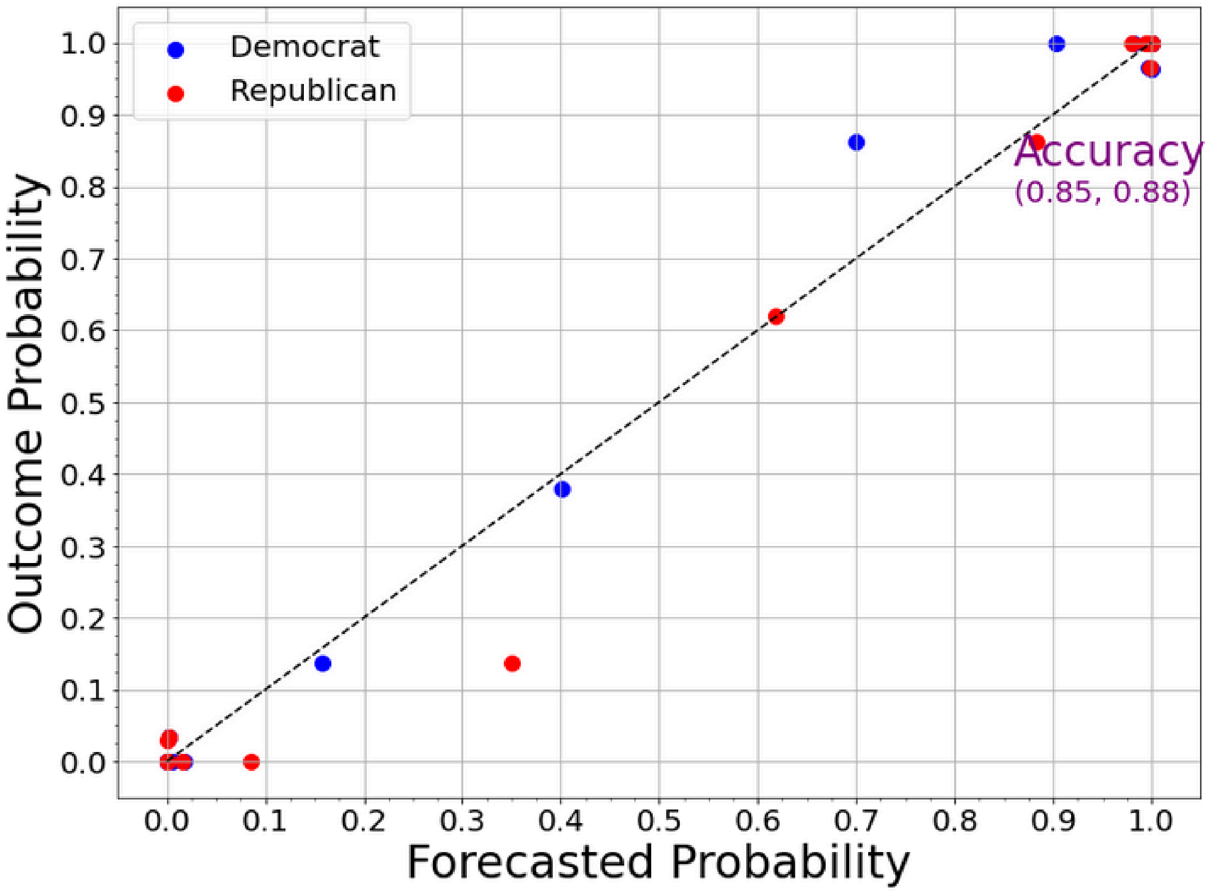


Figure 3: Outcome versus Forecasted Probability of Five-Thirty Eight's US House Election Forecast. Each dot (red - Republican and blue - Democrat) is a histogram bin of the actual win ratio (y-axis) and the average forecast (x-axis). Ideally, all the bins would align along the dashed line and if there was no uncertainty be at the upper right hand corner. The Accuracy (black dot) is a weighted geometric mean along both axes.

FiveThirtyEights last forecast on Monday, November 8th predicted that 227 Republicans and 208 Democrats would win house seats. This did suggest a large Republican victory but they were also clear that there was considerable uncertainty in forecast, specifying an 80% chance that the range would be from a 248-seat Republican win to a 221-seat Democratic win. The final outcome of the House election resulted in 222 Republican seat wins versus 213 Democratic wins. Thus, FiveThirtyEight's forecast was well within is forecasted uncertainty. The percentage of correctly forecasted winners is 96.3%. What is unappreciated about the forecast is that FiveThirtyEight had assigned an average probability of 0.85 to the winning candidates; thus anticipating uncertainty about the final outcome. Let's explore how this average probability is determined and it's accuracy relative to the outcome.

Photrek uses the methods of information theory for its analysis but translates the results to the probability domain to provide an "average forecast". Information theory specifies the average uncertainty of a probability distribution as the arithmetic mean of the logarithm of the probabilities; this is known as the entropy of the distribution. The inverse of the logarithm is the exponential function. When the exponential is applied to the entropy, the result is the geometric mean of the probabilities, providing a measure of the average uncertainty as a probability.

In the plot displayed above, the probabilities forecasted by FiveThirtyEight for the 435 House seats are ordered and grouped into 15 bins for Republicans (red) and Democrats (blue). Each bin contains 29 probabilities. Each bin is assigned a forecast probability (x-axis) and outcome probability (y-axis). The forecast probability for a bin is determined by selecting the probabilities assigned to winning candidates and averaging these probabilities using the geometric mean. The outcome probability for a bin is determined by the ratio of winning candidates in a bin divided by the total number of candidates in a bin. Visually you can see that most of the bins are closely aligned to the diagonal line connection (0,0) and (1,1). This is indicative of high accuracy in the probability forecasts.

A caveat to the analysis is addressing the issue of a zero for either the forecasted or outcome probability. For uncertain events, it is not appropriate to assign a probability of zero, which is reserved for an impossible event. FiveThirtyEight assigns a minimum of 0.00001 to its forecasts, though this is publicly reported as less than 0.01. In a similar vain, we have assigned 0.9999 to winners and 0.0001 to losers for the analysis of the outcome probabilities. If we did not do this, the geometric mean of a group of probabilities containing a zero, would be zero; i.e. the average of a set of events containing an impossible event is the impossible event, just as the arithmetic average of a set of numbers containing infinity is infinity.

From the histogram analysis of the forecasts, one can determine an overall average of both the forecasted probabilities and the outcome probabilities. This then provides a comparison of the forecasters performance against the actual uncertainty of the outcome. In the plot, this is the purple dot specifying the Accuracy. The average along each axis is determined by the geometric mean, weighted by the outcome probability of that bin. In the case of FiveThirtyEight's forecast of the House election, their average forecast is 0.85, which is very close to the average outcome of 0.88. Thus we can say that while their forecast contained uncertainty, they were very accurate in specifying the degree of uncertainty in the forecast.

The cautionary lesson is that forecasters can be very precise in specifying uncertainty of an outcome but this often gets lost as people jump to the headline regarding the forecasted winners and losers. So FiveThirtyEights forecast of 227 Republican wins was inappropriately interpreted as a "Red Wave" without appreciating the effects of the uncertainty specified in this forecast, which turned out to be 85%.

Photrek will publish a more detailed analysis later that includes mathematical specifications of the approach. In the meantime, for those interested in learning more about the method, the book chapter "Reduced Perplexity: A simplified perspective on assessing probabilistic forecasts" in Advances in Info-Metrics: Information and Information Processing across Disciplines provides a tutorial and the journal article "Assessing Probabilistic Inference by Comparing the Generalized Mean of the Model and Source Probabilities" in Entropy provides details.

