2018

Predicting Frac Hits Using Artificial Intelligence; Application in Marcellus Shale

Ryan Tyree

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Tyree, Ryan, "Predicting Frac Hits Using Artificial Intelligence; Application in Marcellus Shale" (2018). *Graduate Theses, Dissertations, and Problem Reports*. 6844.
<https://researchrepository.wvu.edu/etd/6844>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Predicting Frac Hits Using Artificial Intelligence;
Application in Marcellus Shale

Ryan Tyree

Thesis submitted to the Benjamin M. Statler College of Engineering and Mineral
Resources at West Virginia University in partial fulfillment of the requirements for
the degree of

Master of Science
in
Petroleum and Natural Gas Engineering

Shahab D. Mohaghegh, Ph.D., Committee Chair
Samuel Ameri, M.S., Department Chair
Joseph H. Frantz Jr, B.S.

Department of Petroleum and Natural Gas Engineering

Morgantown, West Virginia
2018

Keywords: Frac Hits, Artificial Intelligence, Marcellus
Copyright 2018 Ryan Tyree

Abstract

Predicting Frac Hits Using Artificial Intelligence; Application in Marcellus Shale

Ryan M. Tyree

This work introduces a process of using AI neural networks, for analyzing complex datasets, in order to achieve a higher prediction accuracy in regards to frac hits, at the individual stage level, in the Marcellus Shale when compared to traditional linear methods.

We examined 63 producing wells (parent) along with 79 completed wells (child) to determine the best predictors for accurate frac hit predictions. Our dataset consists of 959 records with 77 predictors and a single binary output (YES or NO) for a frac hit occurrence. Linear methods make analyzing these 77 predictors, along with their interactions, difficult. Neural networks, specifically backpropagation learning algorithm that was used, integrated with a fuzzy pattern recognition algorithm, allow end users to analyze a seemingly endless number of predictors at one time in order to produce a model with increased prediction accuracy over linear approaches. The four techniques discussed include accepting the null hypothesis, a method we refer to as the industry standard, a modified version of the industry standard, and backpropagation algorithms.

In this work we observed a 92.9% prediction accuracy when using a backpropagation neural network. Traditional approaches for the same dataset yield overall accuracies of 73.0%, 64.8%, and 82.8% for the three approaches that are discussed, respectively. Increased prediction accuracy is important because this allows the operator to make proactive data driven decisions for changes in completion design, well spacing, shutting in the parent well prior to the offset frac, or simply doing nothing. These decisions are better justified with increased prediction accuracy, potentially saving the operator valuable time and money.

Acknowledgements

I would like to thank my advisor, Dr. Shahab D. Mohaghegh for his advice, guidance, and patience as I worked on this project. I would also like to thank my committee members Joe Frantz, Samuel Ameri, and Dr. Shahab D. Mohaghegh for their support. This project would not have been possible without the use of the software IMprove™ provided by Intelligent Solutions Inc. and the Appalachian operator that provided the dataset. A special thanks to Matt Curry, and my dad, for never failing to ask me about my progress each and every time we spoke. The consistent reminders were a great motivator to stick with this work and see it to completion. And finally, a big thanks to West Virginia University and the Petroleum and Natural Gas Engineering department for giving me this opportunity.

Table of Contents

Abstract	ii
Acknowledgements	iii
List of Tables	v
List of Figures	vi
Literature Review	1
Background and Problem Definition	6
Development of Dataset	9
Baseline Analytics	15
Backpropagation Neural Network Analytics	23
Conclusions	37
Definition of Terms and Acronyms	39
Cited Literature	41
Appendix A – Definitions of Dataset Columns	42
Appendix B – Backpropagation Neural Network Models	52

List of Tables

Table 1: Results for industry standard practice for predicting frac hits using an x_{CD} of 1500 ft.....	17
Table 2: Results for industry standard practice for predicting frac hits using x_{CD} of 2000 ft.....	17
Table 3: Results for industry standard practice for predicting frac hits using x_{CD} of 2500 ft.....	17
Table 4: Results for modified industry practice for predicting frac hits which includes the predictor of shielded for $x_{CD} = 2500$ ft.....	20
Table 5: Summary table of the prediction accuracies for the Industry and Modified Industry models.	22
Table 6: Summary table for the training, calibration, and verification datasets.....	25
Table 7: Tables of predictors that were included in the original dataset but not used in the predictive models due to these variables not existing until the child well frac is completed.....	26
Table 8: All 35 predictors used as part of the model that includes reservoir properties.	27
Table 9: Performance of the backpropagation neural network model versus the industry practice and modified industry practice prediction techniques.....	30
Table 10: List of predictors used in the final backpropagation model.....	31
Table 11: Prediction accuracy comparison between the two backpropagation models analyzed.....	35
Table 12: Performance of the backpropagation neural network model versus the industry practice and modified industry practice prediction techniques.....	35
Table 13: Overall prediction accuracy comparison of all 5 models/methods investigated.	36
Table 14: Degree of influence for the top four predictors in the final model.....	37

List of Figures

Figure 1: Probability of a well being impacted based on its distance from an infill well in the Woodford shale. (Ajani, A. A., & Kelkar, M. G. 2012, January 1)	2
Figure 2: Conditional probability of a well being impacted Based on its age in the Woodford shale. (Ajani, A. A., & Kelkar, M. G. 2012, January 1)	2
Figure 3: A rate decrease in the producing well, shown in black, occurs at day 72 which corresponds to the same time an offset well, shown in red, was hydraulically stimulated (Lawal, H., Jackson, G., Abolo, N., & Flores, C. 2013, June 10).....	3
Figure 4: A half-length reduction of 50 feet (~15% reduction in area) results in a steeper linear transient flow interpretation when plotted vs. inverse productivity index (Miller et al. 2010). (Lawal, H., Jackson, G., Abolo, N., & Flores, C. 2013, June 10)	4
Figure 5: A reduction of fracture conductivity from 500 md-ft to 0.5 md-ft during a shut-in results in an upward shift in the linear transient flow regime interpretation yielding a higher y-intercept (Lawal, H., Jackson, G., Abolo, N., & Flores, C. 2013, June 10).	4
Figure 6: SPE publications since 2004 showing growing interest in AI within the oil and gas industry. Data current as of 10/13/17.	5
Figure 7: Frac hit illustration	6
Figure 8: Real example showing that the Parent well was only interfered with by a single stage from the child well.	7
Figure 9: Pictorial showing how each of the stages in the child wells were associated to the parent well stages across the studied area.....	10
Figure 10: Pictorial showing how relevant stages were determined.....	11
Figure 11: Percentage of frac stages in a child well that caused a frac hit in an offset parent well with respect to the distance between the wells.	12
Figure 12: Real data example of how frac hits are determined. In this example, stage 14 and 15 from the child well did not have any effect on the parent well.	13
Figure 13: Real data example of how frac hits are determined. In this example all stages beyond stage 6 in the child well caused a frac hit in the parent well.	13
Figure 14: Real data example of how frac hits are determined. In this example all stages beyond stage 4, with the exception of stage 5, in the child well caused a frac hit in the parent well.	14
Figure 15: Prediction accuracy for industry standard practice using x_{CD} of 2500 ft.....	18
Figure 16: Prediction accuracy for modified industry standard practice using x_{CD} of 2500 ft.	21

Figure 17: Backpropagation Neural Network Model	23
Figure 18: Sigmoid Activation Function (https://towardsdatascience.com)	24
Figure 19: Actual frac hit effects compared to the backpropagation neural network outputs for the model that includes reservoir properties as predictors. Total record count is on the X axis and outcome on the Y axis.	27
Figure 20: Calibration and Verification blind test using the weights generated from the training data for the model containing reservoir properties. Total record count is on the X axis and outcome on the Y axis.	28
Figure 21: Prediction accuracy for the backpropagation neural network model that includes reservoir properties.	29
Figure 22: Actual frac hit effects compared to the backpropagation neural network outputs for the final model. Total record count is on the X axis and outcome on the Y axis.	31
Figure 23: Calibration and Verification blind test using the weights generated from the training data for the final model. Total record count is on the X axis and outcome on the Y axis.	32
Figure 24: Prediction accuracy for the backpropagation neural network model that does not include reservoir properties.	33
Figure 25: Comparison of the non reservoir property and reservoir property predictive models. Number of child to parent associations is on the X axis and outcome on the Y axis.	34
Figure 26: Classification of Stage Location	42
Figure 27: Langmuir Volume (www.fekete.com)	45
Figure 28: Illustration showing how the "OnPlane" data point is derived.	46
Figure 29: Illustration explaining how the "Shielded" data point is derived	47
Figure 30: Definition for the output of Effects.	48
Figure 31: Complete list of columns collected as part of the data mining process. The predictors used in the final model are highlighted in blue.	52
Figure 32: Variation of the final model used that contains reservoir properties. This model predicted frac hits with more confidence but ultimately resulted in no additional accuracy.	53
Figure 33: Training data and Calibration/Verification dataset accuracies for the backpropagation neural network model with reservoir properties.	54

Figure 34: Training data and Calibration/Verification dataset accuracies for the backpropagation neural network model that does not contain reservoir properties.	55
---	-----------

Literature Review

Hydraulic fracturing of oil and gas wells began in 1947 on a gas well operated by Pan American Petroleum Corporation in the Hugoton Field (Gidley, J.L., Holditch, S.A., Nierode, D.E. et al. 1989). As of 2015 the EIA had estimated the number of hydraulically fractured oil and gas wells in the US to exceed 300,000 wells. In North America's most active shale fields, the drilling and hydraulic fracturing of new wells is directly placing older adjacent wells at risk of suffering a premature decline in oil and gas production (Jacobs, T. 2017, April 1). The underlying issue has been coined as a "frac hit." And though they have long been a known side effect of hydraulic fracturing, frac hits have never mattered or occurred as much as they have recently, according to several shale experts who say the main culprit is infill drilling (Jacobs, T. 2017, April 1).

Paryani et al noted in their Wolfcamp shale case study that with the increase in drilling infill wells offsetting existing producers, and stacking wells on top of existing producing wells, frac hits and other problems resulting from pressure depletion are becoming a major issue impairing the full field development of unconventional reservoirs (Paryani, M., Smaoui, R., Poludasu, S., Attia, B., Umholtz, N., Ahmed, I., & Ouenes, A. 2017, February 15). In many shale plays, high density wells are thought to be a requirement to effectively drain the resource; thus as well densities increase, frac hits will also increase in occurrence (Lawal, H., Jackson, G., Abolo, N., & Flores, C. 2013, June 10). "It is a very common occurrence—almost to the point where it is a routinely expected part of the operations," said Bob Barree, an industry consultant and president of Colorado-based petroleum engineering firm Barree & Associates. He added that frac hits are also an expensive problem that involve costly downtime to prepare for, remediation efforts after the fact, and lost productivity in the older wells on a pad site (Jacobs, T. 2017, April 1).

Several factors are considered to be drivers behind frac hits. These include, but are not limited to: fracture plane anisotropy, well spacing (distance between laterals), natural fracturing, and depleted

zones (changing stress fields); all of which are examined in this work. In this research, 91% of all frac hits observed were in the northeast/southwest direction, which is consistent with the known direction of hydraulic fractures in the studied formation. Ajani & Kelkar found that as the distance between wells increases, the probability of interference decreases as shown in **Figure 1** (Ajani, A. A., & Kelkar, M. G. 2012, January 1).

Figure 1: Probability of a well being impacted based on its distance from an infill well in the Woodford shale. (Ajani, A. A., & Kelkar, M. G. 2012, January 1)

Ajani & Kelkar also noted (**Figure 2**) that with increased depletion from the producing well the probability of interference increases because the frac fluid preferentially enters zones of pressure depletion.

Figure 2: Conditional probability of a well being impacted Based on its age in the Woodford shale. (Ajani, A. A., & Kelkar, M. G. 2012, January 1)

As the wells produce reservoir fluids (gas and water in the case of the Haynesville shale), their reservoir pressure decreases as a function of production drawdown, creating zones of pressure depletion (Esquivel, R., & Blasingame, T. A. 2017, July 24). Esquivel et al also stated that these pressure sinks attract influxes from fluids and proppant that are being injected in offset locations. The fracture initiated from an offset well close to this drained envelope is preferentially extended towards the depleted zone due to reduced closure stress in that direction (Mukherjee, H., Poe, B. D., Heidt, J. H., Watson, T. B., & Barree, R. D. 2000, August 1). Mukherjee et al went on to state that “hydraulic fracture geometry evolves asymmetric wings when induced into pressure-depleted zones”; suggesting that the fracture geometry in the child well could be negatively impacted from the frac hit. Esquivel et al mentioned that while out of the scope of their particular study, reservoir and geomechanical properties could also play a role in the occurrence and magnitude of a given frac-hit. Both reservoir and geomechanical properties, or analogs for each, are considered in this research.

Lawal et al analyzed several examples from the Haynesville and Marcellus shale and found that inference (*from a frac hit*) has been known to typically reduce (occasionally enhance) the performance of wells currently in production by altering the existing fracture network, or near wellbore permeability, via the presence of multiple phases.

Figure 3: A rate decrease in the producing well, shown in black, occurs at day 72 which corresponds to the same time an offset well, shown in red, was hydraulically stimulated (Lawal, H., Jackson, G., Abolo, N., & Flores, C. 2013, June 10).

Lawal et al displayed graphically the concept of reduction in fracture area and a reduction in fracture conductivity resulting from fluid invasion from a child well.

Figure 4: A half-length reduction of 50 feet (~15% reduction in area) results in a steeper linear transient flow interpretation when plotted vs. inverse productivity index (Miller et al. 2010). (Lawal, H., Jackson, G., Abolo, N., & Flores, C. 2013, June 10)

Lawal et al brought attention to the concept of the y-intercept changing when the fracture conductivity changes but does not change when the fracture area reduces.

Figure 5: A reduction of fracture conductivity from 500 md-ft to 0.5 md-ft during a shut-in results in an upward shift in the linear transient flow regime interpretation yielding a higher y-intercept (Lawal, H., Jackson, G., Abolo, N., & Flores, C. 2013, June 10).

Well to well interference, and the effects of this interference, is widely documented in literature. However, methods used to predict the occurrence(s) of frac hits is scarcely discussed. When viewed from a holistic point, frac hits in all but the most homogenous formations may not be preventable, even by reasonable increases of distance between wells or equalizing fracture injection volumes (King, G. E., Rainbolt, M. F., & Swanson, C. (2017, October 9). Ajani & Kelkar studied the age of the producing well, as an analog to pressure depletion, as well as the distance between wells, but did not discuss a method for analyzing the interactions between the two. These correlations are where the use of artificial intelligence (AI) becomes useful in solving complex analytical problems utilizing large datasets. AI techniques have been used to map the natural fracture network in the Utica shale (Mohaghegh 2017), prediction of bubble point pressure (Alakbari et al 2016), and predicting wellbore instability (Okpo et al 2016); to name a few. AI has become increasingly popular in the oil and gas industry over the past decade but has not yet breached the idea of more accurately predicting the occurrence of frac hits. **Figure 6** illustrates the growing popularity of AI in the oil and gas industry.

Figure 6: SPE publications since 2004 showing growing interest in AI within the oil and gas industry. Data current as of 10/13/17.

This thesis contains data and analyses that, as far as the researchers are aware, are a first of their kind in regards to frac hits for the oil and gas industry. We demonstrate the ability to more accurately predict the occurrence of frac hits using AI, more specifically the method of backpropagation neural networks.

Background and Problem Definition

What is a fracture “frac” hit? A frac hit refers to the act of a producing well (referred to as the parent well hereafter) being interfered with by the fracture treatment of a nearby offset well (referred to as the child well hereafter). In most literature, the above reference is made in a very broad sense by saying that well A hit well B, Well C hit well D, and so on. In horizontal wells this summarization is not entirely representative of reality, due to the nature of how horizontal wells are drilled and completed as shown in **Figure 7**:

Figure 7: Frac hit illustration

The generalization of the child well hitting the parent well holds true, but the hit(s) occur from multiple stages initiating from the child well. This granularity is important when studying frac hits as, it was found in this research, not all stages from the child well will interfere with the parent well.

Figure 8: Real example showing that the Parent well was only interfered with by a single stage from the child well.

The sharp increase in casing pressure is indicative of a frac hit as shown in **Figure 8**. But why did only one of the ten stages shown in this illustration cause interference in the parent well? This is the question we set out to answer with this research. By developing a better understanding of the per stage interference we, as an industry, will be better positioned to make data driven decisions on completion techniques, landing targets, well spacing, etc. in order to reduce or eliminate the occurrence of frac hits. More real data examples can be found in **Figure 12**, **Figure 13**, and **Figure 14**.

In our literature review we found mention of stress anisotropy, natural fracture networks, distance between wells, and reservoir depletion as common attributors to frac hits. A complete list of our predictors, along with definitions, is included in **Appendix A – Definitions of Dataset Columns**. The goal of this research is to answer the following question: “Is it possible to use artificial intelligence to increase accuracy in predicting frac hits, on a per stage basis, when compared to methods currently utilized by the oil and gas industry?” Three methods, all using linear equations to predict frac hits, will be discussed. The methods consist of accepting the null hypothesis, a method we refer to as the current

industry standard, and a modification to the industry standard based on findings from this research. The industry standard approach was determined by polling a number of Appalachia operators and deriving a simple mathematical approach to illustrate the decision-making process that consists of individuals looking at maps and basing their conclusions on past experiences. The results of these three methods are then compared to the results of the backpropagation neural networks to determine the success of project.

Development of Dataset

The final dataset contains 77 columns (complete list and descriptions located in **Appendix A – Definitions of Dataset Columns**), known as predictors, and a single outcome collected from 63 producing (parent) wells and 79 completed (child) wells which contained a total of 979 individual frac stages located within a 7.5 by 15 square mile area. The predictors consisted of numerous binary variables as well as real numbers. The original goal of this work was to predict the impact of a frac hit on a parent well by developing correlations between the predictors and the casing pressure responses associated with the frac hits; then developing correlations between the casing pressure responses and the realized production results. This proved to be difficult considering that each casing pressure response in the parent well had a corresponding frac stage in the child well, but the overall production response in the parent well is the cumulative result of the total number of frac hits that occurred. Ultimately, the original goal of this work was determined to exceed the scope of this project. Tools such as production logs, DAS (Distributed Acoustic Sensing), and DTS (Distributed Temperature Sensing) would be valuable datasets to aid in identifying the production responses for individual stages prior to a frac hit and after a frac hit occurrence. This type and granularity of data was unavailable to the researchers at the time this work was completed.

Results from the impact models, previously mentioned, indicated that using neural networks to predict the occurrence of a frac hit in the parent well may prove to be successful. This effort removed the production response variable from the analysis by focusing only on modelling individual frac hits to a parent well. By doing so, the dataset contains a known outcome of a frac hit (Yes-1 or No-0) in the parent wells for each of the child well frac stages. The idea of predicting the likelihood of a frac hit (at the stage level) in a parent well, prior to completing the child well, is advantageous to the operator as this gives the operator the means of adjusting the completion design on a per stage basis in the child well in order to mitigate, or eliminate, frac hits.

Several assumptions were made, and pre analytic data mining completed, to prepare the dataset for analysis using artificial intelligence. These assumptions need to be clearly discussed before moving on to the data analytics. A starting point needed to be determined in order to begin building the dataset. We first began by taking the stages within our studied area and measuring the distance between each stage and all other stages in the dataset, as long as the relationship between the two wells was that of a child and parent. As such, the oldest producing well in the studied area was a parent well to all others and the newest producing well was a child well to all others with every other well falling somewhere in the middle.

Child Well Stage to Parent Well Stage Associations

Figure 9: Pictorial showing how each of the stages in the child wells were associated to the parent well stages across the studied area.

This process resulted in a very large and confusing data matrix, containing almost 3.6 million rows of data, with distances between wells of up to several miles. The frac hit effect, or output variable, had to be manually interpreted for each and every stage which will be discussed later in this section. Knowing

this, a mechanism for filtering the large data matrix needed to be developed. This was completed in two steps. First, we reduced the associations to stages with the minimum distance between child and parent wells so that each stage in a child well could not be associated to more than one stage in each parent well, as seen in **Figure 10**, by making the assumption that the impacted stage of the parent well was most likely the stage that was in closest proximity to the child well stage. This is a major assumption in this research and one that had to be made in order to develop a method to link the various parameters in each of the wells to one another. In reality, within the confines of the studied dataset, we have no real way of knowing at which point along the length of the lateral that the parent well was impacted from the child well frac stage without the use of sophisticated technologies such as DAS or DTS. The result of the filtering mechanism describe above is represented in **Figure 10**.

Figure 10: Pictorial showing how relevant stages were determined.

The second step used to filter the dataset even further was to evaluate the percentage of hits that occur with respect to the distance between wells. The first filtering step only condensed the data

matrix to list what we referred to as “relevant stages”. These relevant stages still contained associations with distances exceeding several miles. Based on past experiences it is known that frac hit typically do not occur at these distances. **Figure 11** shows the percentages of frac stages that result in a frac hit with respect to the distance between the child and parent wells.

Figure 11: Percentage of frac stages in a child well that caused a frac hit in an offset parent well with respect to the distance between the wells.

It can be determined that the frequency of frac stages resulting in a frac hit decrease as the distance between wells increase. We chose to use 2,500 ft as our cutoff for the dataset based on the consistently low percentage of 10% for frac hit occurrence at distances greater than 2,000 ft. The two steps described above reduced the initial dataset of 3,588,933 child to parent well associations down to 959. It was chosen to take this overall approach in developing the dataset as this allowed the final dataset to remain free of user bias.

The effects, or output, was determined manually by reviewing each child well frac stage and how the stage impacted the offset parent well. Casing pressure inflections were used as the

determining factor for deciding if a frac hit occurred, or did not occur. The following figures are examples from the dataset used to show how the process of picking the frac hits was completed.

Frac Hit Example 1 (Field Data) – Effects Output

Figure 12: Real data example of how frac hits are determined. In this example, stage 14 and 15 from the child well did not have any effect on the parent well.

Frac Hit Example 2 (Field Data) – Effects Output

Figure 13: Real data example of how frac hits are determined. In this example all stages beyond stage 6 in the child well caused a frac hit in the parent well.

Frac Hit Example 3 (Field Data) – Effects Output

Figure 14: Real data example of how frac hits are determined. In this example all stages beyond stage 4, with the exception of stage 5, in the child well caused a frac hit in the parent well.

Stages labeled in **red** above for *Figure 12*, *Figure 13*, and *Figure 14* represent a frac stage that caused a frac hit in the parent well. The frac hit is determined by an inflection in the casing pressure. Stages labeled in **black** did not result in a frac hit. These graphs were generated for each child to parent well relationship and every stage evaluated by the researcher to determine the occurrence of a frac hit. Ultimately it was determined that 259 of the 959 stages resulted in a frac hit, while the remaining 700 did not.

The final step in developing the dataset was joining the 77 data columns used as the predictors in the neural network analysis. These predictors consist of completion parameters, reservoir properties, stratigraphy, physical location, depths, etc. This data was pulled from several different databases and software sources at the stage level and joined to the dataset based on the well and stage the data belonged too. The end result was a complete dataset with predictors unique to the stages being evaluated. A complete list of these data columns is listed, along with definitions for each, in **Appendix A – Definitions of Dataset Columns**.

Baseline Analytics

As mentioned previously, the goal of this research was to increase the prediction accuracy of frac hits. Three methods were used to develop baselines to which the neural network results were compared against. The first process, known as accepting the null hypothesis, was used as a basis to compare all other methods in order to determine effectiveness. Linear processes have been used for a number of years in determining frac hits within the oil and gas industry. This typically consists of an engineer looking at a map displaying a parent well and planned child well, judging the distance between wells based on well spacing in the field, and (based on past experience) making a determination on whether or not the parent well will be hit by the offset child. A second linear process was developed from this work as a means of challenging the neural networks, which also proved to be inferior. The thought processes mentioned above can be explained in mathematical form, which is detailed in the following sections.

The first process we will discuss is a statistical concept known as accepting the null hypothesis. This theory states that the null hypothesis assumes that whatever you are trying to prove (in this case a frac hit occurred) did not actually happen. For the purpose of this research, accepting the null hypothesis is the equivalent of classifying each row in the dataset as no frac hit occurring. By doing so, we accurately predict 700 of the 959 records, or 73% accuracy. This prediction accuracy is used as the baseline to compare against all other prediction methods. In theory, other prediction models should exceed 73% accuracy to be considered viable.

The second process, which I will refer to as industry practice (this practice was determined to be industry practice based on discussions held with a number of Appalachia operators), uses the horizontal distance between wells and the known frac plane to determine frac hits. This can be explained mathematically as:

$$y = x < x_{CD} \rightarrow \text{True} = 1; \text{False} = 0$$

$$\text{On Plane} \rightarrow \text{Yes} = 1; \text{No} = 0$$

Where: x = known distance between child well and parent well

x_{CD} ¹ = user defined “cutoff distance” that is believed to be the maximum reasonable distance that a frac hit can occur based on an individual’s experience in the production field

On Plane = The parent well is within the frac plane of the stage being completed on the child well. See “OnPlane” in Appendix A – Definitions of Dataset Columns for a complete definition.

From these two equations a decision matrix is created with **red** representing positive frac hit prediction:

	-	
	TRUE	FALSE
Yes	1,1	1,0
No	0,1	0,0

...where 1,1 = frac hit → True, all others NO hit

Example:

x_{CD} → Cutoff Distance = 1500 ft (user defined)

x → Distance = 1300 ft

On Plane = Yes

Solution:

$$y = 1300 < 1500 = 1$$

$$\text{On Plane} = 1$$

Decision matrix predicts that frac hit will occur.

¹ x_{CD} is a function of formation thickness and general understanding of the natural fracture network. Knowledge and experience in the field are keys to determining this distance and will vary dependent on the user.

When this methodology was applied to all child well to parent well associations within the dataset the results shown in **Table 1**, **Table 2**, and **Table 3** are achieved using x_{CD} of 1500 ft, 2000 ft, and 2500 ft:

Industry Standard Prediction Method ($x_{CD} = 1500$)

Distance	Stage Count	Stage Count NO HIT	Stage Count YES HIT	Correct NO HIT Predictions	Correct YES HIT Predictions	Total Correct Predictions
0 - 1500	249	121	128	61	124	185
1501 - 2000	304	212	92	212	0	212
2001 - 2500	406	367	39	367	0	367

Table 1: Results for industry standard practice for predicting frac hits using an x_{CD} of 1500 ft.

Industry Standard Prediction Method ($x_{CD} = 2000$)

Distance	Stage Count	Stage Count NO HIT	Stage Count YES HIT	Correct NO HIT Predictions	Correct YES HIT Predictions	Total Correct Predictions
0 - 1500	249	121	128	61	124	185
1501 - 2000	304	212	92	126	89	215
2001 - 2500	406	367	39	367	0	367

Table 2: Results for industry standard practice for predicting frac hits using x_{CD} of 2000 ft.

Industry Standard Prediction Method ($x_{CD} = 2500$)

Distance	Stage Count	Stage Count NO HIT	Stage Count YES HIT	Correct NO HIT Predictions	Correct YES HIT Predictions	Total Correct Predictions
0 - 1500	249	121	128	61	124	185
1501 - 2000	304	212	92	126	89	215
2001 - 2500	406	367	39	191	30	221

Table 3: Results for industry standard practice for predicting frac hits using x_{CD} of 2500 ft.

Several different values of x_{CD} are used because this is a user defined variable and will be dependent on the individual performing the evaluation and their experiences. For the purpose of this research, we will focus on x_{CD} of 2500 ft since this was the cutoff that we used for the dataset. With respect to the industry practice for predicting frac hits, using an x_{CD} of 2500 ft will classify all rows as a frac hit so long as the child well stage and parent well stage are on plane with one another. Graphical representation for the results using $x_{CD} = 2500$ is shown in **Figure 15**.

Figure 15: Prediction accuracy for industry standard practice using x_{CD} of 2500 ft.

To discuss **Figure 15** further, at first glance it would appear that the industry practice does an excellent job at predicting when a frac hit will occur with a prediction accuracy between 90% - 100% for any distance less than or equal to 2000 ft. The total accuracy for predicting a YES over this span is 96.7% which appears to be fantastic. However, over the same span, the prediction accuracy ranges from 15% - 80% for predicting when a hit will not occur with an average prediction accuracy of just 55.1%. The prediction accuracy for the entire dataset was 64.75%, accurately predicting 621 of the 959 records, making the industry standard practice a worse model than merely accepting the null hypothesis. The

industry practice model is biased towards predicting that a frac hit will always occur so long as the child well stage and parent well stage are on plane. This bias could lead an operator to spend unnecessary money attempting to prevent an incident that may never occur or defer production due to a preemptive shut-in that did not need to take place. Therefore, we feel that it is important to predict both NO hits and YES hits as accurately as possible with a single model to give the end user the tools necessary to make the best data driven decisions.

The third process used to predict frac hits is a modified version of the industry standard practice. The modified version is an outcome of this research based on our findings that the shielded predictor carried tremendous weight in predicting frac hit occurrence. Shielded is a binary variable which made it easy to incorporate into the industry standard model as an additional predictor. This model is mathematically represented as follows:

$$y = x < x_{CD} \rightarrow \text{True} = 1; \text{False} = 0$$

$$\text{On Plane} \rightarrow \text{Yes} = 1; \text{No} = 0$$

$$\text{Shielded} \rightarrow \text{Yes} = 1; \text{No} = 0$$

Where: x = known distance between child well and parent well

x_{CD} = user defined “cutoff distance” that is believed to be the maximum reasonable distance that a frac hit can occur based on experience

On Plane = The parent well is within the frac plane of the stage being completed on the child well

Shielded = The parent well is protected by another producing well in between the child well. See “Shielded” in Appendix A – Definitions of Dataset Columns for a complete definition.

From these two equations a decision tree is generated with **red** representing positive frac hit prediction:

$(x < x_{CD}, \text{On Plane}, \text{Shielded}) = (1,1,1); (1,1,0); (1,0,1); (0,1,1); \dots$ where 1,1,0 = Frac Hit $\rightarrow$ TRUE

Example

$x_{CD} \rightarrow$ Cutoff Distance = 1500 ft (user defined)

$x \rightarrow$ Distance = 1300 ft

On Plane = Yes

Shielded = Yes

Solution

$$y = 1300 < 1500 = 1$$

On Plane = 1

Shielded = 1

Decision tree predicts that frac hit will not occur.

This methodology was applied to all rows within the dataset with the following results being achieved at x_{CD} of 2500 ft:

Modified Industry Standard Prediction Method ($x_{CD} = 2500$)

Distance	Stage Count	Stage Count NO HIT	Stage Count YES HIT	Correct NO HIT Predictions	Correct YES HIT Predictions	Total Correct Predictions
0 - 1500	249	121	128	62	123	185
1501 - 2000	304	212	92	189	63	252
2001 - 2500	406	367	39	347	10	357

Table 4: Results for modified industry practice for predicting frac hits which includes the predictor of shielded for $x_{CD} = 2500$ ft.

The modified industry standard model does not carry the same bias as its predecessor because of the added 3rd dimension. Therefore, the prediction accuracy for the NO hits increases. **Figure 16** graphically shows how the prediction accuracy of the NO hits has increased, but the model sacrifices accuracy at the expense of accurately predicting YES hits for distances beyond 1800 - 2000 ft. This is largely due to the well spacing in the studied field which was drilled on 1000 ft spacing. The modified industry standard takes into account the predictor of shielded, which becomes a factor at approximately 2000 ft in regards to 1000 ft

well spacing. The modified industry practice model is biased towards predicting that no frac hit will occur at distances beyond 2000 ft because, in most instances, there is an existing producing well drilled between the child well being frac'd and the parent well being evaluated. This concept is better represented graphically in **Figure 16**. The accuracy for predicting a YES occurrence in latter distances suffers because the likelihood of a shield existing at these distances is high. See Appendix A – Definitions of Dataset Columns for further explanation on the concept of “Shielded”.

Figure 16: Prediction accuracy for modified industry standard practice using x_{CD} of 2500 ft.

When compared to the industry practice model, the prediction accuracy for predicting a YES at distances less than or equal to 2000 ft is 84.5%, which is less accurate. However, the prediction accuracy for predicting a NO across the same span slightly increases to 66.8%. The modified industry practice model outperforms its predecessor at distances greater than 2000 ft with predicting NO frac hits, but lags when predicting a YES with prediction accuracy averages

of 93.0% vs 52.5% and 37.7% vs 81.1%, respectively. A combination of the two models would be better than either individually, but neither model is very accurate at predicting a NO at distances less than 1500 ft. At these distances both models are strongly biased towards predicting a YES because of the close proximity of the child and parent as well as the lack of a shield. Both models share the same prediction accuracy of just 41.0% with extreme bias towards classifying all instances that are on plane as a YES. **Table 5** below summarizes the prediction accuracies across the distance ranges for both YES and NO outcomes using both industry models.

Distance	Industry Practice Model		Modified Industry Practice Model	
	NO	YES	NO	YES
0 - 1500	48.1%	96.4%	48.8%	95.8%
1501 - 2000	66.8%	96.4%	88.4%	74.7%
2001 - 2500	52.5%	81.1%	93.0%	37.7%

Table 5: Summary table of the prediction accuracies for the Industry and Modified Industry models.

Backpropagation Neural Network Analytics

Backpropagation neural networks are a form of supervised machine learning that were used in this research as the tool to remove all user bias from the analytics. The software IMprove™, provided by Intelligent Solutions, Inc.

(<http://www.intelligentsolutionsinc.com/index.shtml>), was used to compute the backpropagation algorithms against the previously discussed dataset. These results are detailed in the following sections, but first a general understanding of how backpropagation neural network models function is discussed. Detailed calculations are not discussed as part of this research but are well defined in literature if further interest is desired. **Figure 17** below illustrates the workflow behind the backpropagation algorithm.

Figure 17: Backpropagation Neural Network Model

A simplistic step by step process is detailed in the following bullet points:

1. Multiply the *Inputs* by the individual weights in the *Weight Matrix* (represented by the arrows).
2. Add these multiplications together at each *Hidden Layer* node.
3. Multiply the *Bias* by its weight in the *Weight Matrix* and add this to the value in the *Hidden Layer*.
4. Apply an *Activation Function* to the *Hidden Layer* (variable z in **Figure 18**) to adjust the calculated *Hidden Layer* variable in Step 3 to derive a value between 0 and 1.

Figure 18: Sigmoid Activation Function (<https://towardsdatascience.com>)

5. The *Hidden Layer* now acts like the *Input layer* and the Steps 1-4 are repeated against the *Hidden Layer* to calculate the *Output(s)*. This research example shows a single *Output* in the model. However, backpropagation neural networks are capable of handling multiple outputs.

****Note:** The process detailed in steps 1-5 is referred to as forward propagation.

The objective of a backpropagation neural network is to reduce the RMSE (Root Mean Square Error) to the absolute minimum possible value. This is accomplished by working backwards through the network, adjusting the weights in order to recalculate the *Hidden Layer* and ultimately calculate a new *Output*. The entire process of forward propagation combined with backwards propagation to calculate a new *Output* is known as an epoch. Numerous

epochs are performed against the training set and RMSE calculated for each epoch. The algorithms adjust the weights slightly with each epoch with the goal of reducing the RMSE with each additional epoch. The weights derived within each epoch are then used to calculate the RMSE for the calibration and verification datasets as a blind and double-blind test for determining the viability of each model. The training, calibration, and verification subgroups are split up as 80%, 10%, and 10% of the data, respectively. A breakdown of the known outcomes included in these subgroups is shown in **Table 6**.

Distance	Training		Calibration		Verification		
	NO	YES	NO	YES	NO	YES	
0 - 1500	98	103	11	12	12	13	
1501 - 2000	173	72	16	9	23	11	
2001 - 2500	289	32	43	5	35	2	
	560	207	70	26	70	26	959

Table 6: Summary table for the training, calibration, and verification datasets.

IMprove™ is capable of computing an output using all 77 predictors (complete list located in **Appendix A – Definitions of Dataset Columns**) as inputs into the network. In theory, the process of backpropagation will adjust the weights to favor the more significant variables and minimize the impact of the less significant predictors to reduce the RMSE in the model. However, the computing power needed for these computations is significant and therefore model run times can be time extensive. A technique such as fuzzy pattern recognition is used to determine a first order degree of influence to develop a tornado chart showing the significance of each predictor, relative to the next, which provides a starting point for input selection into the backpropagation neural network models. Fuzzy pattern recognition was used against the dataset in this research as a starting point for the backpropagation inputs. A summary of these influences is shown in the countless models were run, using a number of

different predictors in each, with two models being used as the final product for discussion in this paper.

The two models discussed in this section share many of the same predictors with one model including reservoir properties (**Table 8** on page 27) for the child and parent wells while the other model does not (**Table 10** page 31). This research shows that the model containing reservoir properties is able to predict frac hits with a higher level of confidence, but the overall accuracy of predicting a YES or a NO is roughly the same. Both models show an increase in prediction accuracy over previously discussed methods. The decision on which model is preferred is left to the user, but this research indicates that a less complicated model can be just as accurate.

Goals were established prior to building the models within IMprove™. The objective of this research is to increase the prediction accuracy of frac hits to allow oil and gas producers to make data driven decisions in regards to operational changes for frac hit prevention and/or mitigation. Upon examining the dataset used in this research, one will find that a number of the predictors associated with the child well (designation FW for Frac Well) are not known variables until the frac is completed. For this reason, these columns are not desirable for inclusion in the neural network models because the goal is to predict frac hits prior to beginning the completion, at a time when these data variables do not yet exist. **Table 7** below contains a list of 12 child well variables that were not used in the predictive models.

FW Frac Gradient	FW BrkD Frac Rate BPM
FW Max Frac Pressure	FW_Total_Sand lbs
FW Avg Frac Pressure	FW_%_Fine_Sand
FW BrkD Frac Pressure	FW_CLNVol bbls
FW Max Frac Rate BPM	FW_SLRYVol bbls
FW Avg Frac Rate BPM	FW_Max Prop Conc PPG

Table 7: Tables of predictors that were included in the original dataset but not used in the predictive models due to these variables not existing until the child well frac is completed.

As mentioned previously, a number of iterations were run, on many combinations of variables, as an attempt to remove user bias and exhaust all possible solutions. Two models were ultimately decided on with the results of those models explained in detail in the following sections. The first model examined includes the 35 predictors, shown in **Table 8**, used within the model that includes reservoir properties. The HW designation represents the Parent well.

Location	FW_VL_Marcellus	HW_CLNVol bbls
FW_UTM X	Delta_Days_HW Producing Before Hit	HW_SLRYVol bbls
FW_UTM Y	MIN3D_Distance	HW_NetH_TOTAL
FW_TVDSS	TVD_Difference	HW_Total_TOC
FW_NetH_TOTAL	OnPlane	HW_Total_DEN
FW_Total_TOC	Shielded	HW_Total_PHIE
FW_Total_DEN	HW_UTM X	HW_Total_PERM
FW_Total_PHIE	HW_UTM Y	HW_Total_SW
FW_Total_PERM	HW_TVDSS	HW Ro Marcellus
FW_Total_SW	HW BrkD Frac Rate BPM	HW_PL_Marcellus
FW Ro Marcellus	HW_Total_Sand lbs	HW_VL_Marcellus
FW_PL_Marcellus	HW_%_Fine_Sand	

Table 8: All 35 predictors used as part of the model that includes reservoir properties.

Processing the predictors in Table 8 through a backpropagation neural network yields the following results shown in **Figure 19** for all 959 records in the dataset.

Figure 19: Actual frac hit effects compared to the backpropagation neural network outputs for the model that includes reservoir properties as predictors. Total record count is on the X axis and outcome on the Y axis.

As seen in **Figure 19**, the actual effects in the dataset are represented by green dots and the predicted effects, or output, are represented by red dots. The shaded area is a representation of the model's confidence for accurately predicting the output. For reference, a perfect predictive model would display no shaded area. This will be examined further later in this section when the reservoir property model is compared to the final model that does not include reservoir properties.

Figure 20 below demonstrates that the weights calculated from the training data are capable of accurately predicting outputs from data that was not used as part of the training set. This type of test is necessary when using neural networks because it is possible for the models to “learn” the training data and therefore may be subpar at accurately predicting data that was not included in the training phase. This is not optimal as the goal of these models is to process data not yet seen by the neural network to assist the operator with understanding which stages in the child wells will result in a frac hit on the offset producing parent wells.

Figure 20: Calibration and Verification blind test using the weights generated from the training data for the model containing reservoir properties. Total record count is on the X axis and outcome on the Y axis.

Figure 19 and **Figure 20** show that the backpropagation neural network model, unlike the industry practice and modified industry practice model, predicts the output as a real number between zero and one. In order to perform a fair comparison, we must convert the real number solution into a binary one. To do so, the value of 0.5 is used to assign each of the backpropagation neural network predictions as a NO, or a YES, respectively.

$$x < 0.5 = 0, \quad x > 0.5 = 1$$

The converted binary predictions of are then compared to the known outputs to determine the accuracy of the model. **Figure 21** shows the performance of the backpropagation neural network model with reservoir properties.

Figure 21: Prediction accuracy for the backpropagation neural network model that includes reservoir properties.

Table 9 details the performance of the backpropagation neural network model with reservoir properties versus the industry practice techniques outlined in the **Baseline Analytics** section. The neural network model vastly outperforms the industry practice techniques (looking at both NO and YES combined) for distances less than 2000 ft, but is moderately better than the industry models at

distances greater than 2000 ft.

Distance	Industry Practice Model		Modified Industry Practice Model		Backpropagation Nnet with Res Properties	
	NO	YES	NO	YES	NO	YES
0 - 1500	48.1%	96.4%	48.8%	95.8%	88.3%	94.2%
1501 - 2000	66.8%	96.4%	88.4%	74.7%	95.3%	83.6%
2001 - 2500	52.5%	81.1%	93.0%	37.7%	98.5%	41.9%

Table 9: Performance of the backpropagation neural network model versus the industry practice and modified industry practice prediction techniques.

Referring to **Table 6** on page 25, the record count for YES decreases with each incremental increase in distance between the parent and child wells. As such, the count for YES at distances greater than 2000 ft is only 32 records in the training set, which is less than half and one third of the preceding two distance ranges. It is important to note that neural networks are a supervised machine learning technique that do exactly that; learn. Much like humans, we require textbooks, homework assignments, and repetition to learn a new subject. Supervised machine learning is no different, except the human is replaced by the computer. These techniques require enough data to “learn” a new subject, and in the event where not enough data is provided, it is possible that the desired outcome will not be achieved. In summary, it is likely that the model was not provided enough data for the YES designation at distances greater than 2000 ft to develop accurate predictions at these distances. The same holds true for the backpropagation neural network model without reservoir properties which is discussed in the following section.

The final model discussed in this paper is similar to the previous neural network discussed, except reservoir properties are excluded. **Table 10** below is a complete list of the columns used in the final model. As seen, reservoir properties for both the parent well and child well were excluded. The total number of predictors was cut from 35 down to 17.

Location	HW_UTM X
FW_UTM X	HW_UTM Y
FW_UTM Y	HW_TVDSS
FW_TVDSS	HW BrkD Frac Rate BPM
Delta_Days_HW Producing Before Hit	HW_Total_Sand lbs
MIN3D_Distance	HW_%_Fine_Sand
TVD_Difference	HW_CLNVol bbls
OnPlane	HW_SLRYVol bbls
Shielded	

Table 10: List of predictors used in the final backpropagation model.

The results for the final model strongly resemble those of the previous model that included reservoir properties as shown in **Figure 19**. **Figure 22** shows the output for all records in the final model.

Figure 22: Actual frac hit effects compared to the backpropagation neural network outputs for the final model. Total record count is on the X axis and outcome on the Y axis.

Results for the final model's calibration and verification datasets are similar as well as shown in **Figure 23**. However, it should be noted that the non-reservoir property model inaccurately predicted four NO effects as YES, with an output of 1, in the blind and double blind tests (signifying a very high confidence), while the model that included reservoir properties, shown in **Figure 20**, did not.

Figure 23: Calibration and Verification blind test using the weights generated from the training data for the final model. Total record count is on the X axis and outcome on the Y axis.

The performance of the backpropagation neural network model without reservoir properties is shown in **Figure 24**. The backpropagation neural network without reservoir properties breaks down at predicting a YES at a very similar distance to the model with reservoir properties. It is again concluded that the training dataset, of just 32 records for YES, is not substantial enough to teach the model at distances greater than 2000 ft.

Figure 24: Prediction accuracy for the backpropagation neural network model that does not include reservoir properties.

Looking at the model comparisons another way, the reservoir property model predicts outcomes with higher confidence when compared to the model without reservoir properties. **Figure 25** shows the modeled outputs, represented by red and blue dots, against the known effects shown in green. For reference, a perfect model match would trace the green line. Therefore, it can be said that the reservoir property model is a closer match to the known outcomes. However, it should not be concluded that the model is more accurate, only more confident.

Figure 25: Comparison of the non reservoir property and reservoir property predictive models. Number of child to parent associations is on the X axis and outcome on the Y axis.

Represented numerically, **Table 11** shows the backpropagation neural network model without reservoir properties compares well to the neural network model containing reservoir properties, with a slight edge in predicting a YES at distances greater than 2000.

Distance	Neural Network Model with Res Properties		Neural Network Model without Res Properties	
	NO	YES	NO	YES
0 - 1500	88.3%	94.2%	90.6%	90.1%
1501 - 2000	95.3%	83.6%	94.1%	86.2%
2001 - 2500	98.5%	41.9%	99.3%	51.2%

Table 11: Prediction accuracy comparison between the two backpropagation models analyzed.

Both models compare well with one another once converted into binary predictions, as shown in **Table 11**, with the model lacking reservoir properties edging out the competing model at distances greater than 2000 ft. **Table 12** provides a final comparison of all the models discussed (excluding null hypothesis) across the three distance ranges.

Distance	Industry Practice Model		Modified Industry Practice Model		Backpropagation Nnet with Res Properties		Backpropagation Nnet without Res Properties	
	NO	YES	NO	YES	NO	YES	NO	YES
0 - 1500	48.1%	96.4%	48.8%	95.8%	88.3%	94.2%	90.6%	90.1%
1501 - 2000	66.8%	96.4%	88.4%	74.7%	95.3%	83.6%	94.1%	86.2%
2001 - 2500	52.5%	81.1%	93.0%	37.7%	98.5%	41.9%	99.3%	51.2%

Table 12: Performance of the backpropagation neural network model versus the industry practice and modified industry practice prediction techniques.

Referring back to previous comments, the neural network model with reservoir properties predicts outcomes with a higher confidence, but this does not necessarily lead to increased accuracies. Once the modeled outcomes are converted to binary predictions the neural network model without reservoir properties is slightly more accurate in both NO and YES predictions. This comparison, as well as overall accuracy comparisons to all previously discussed analytical methods, is shown in **Table 13**.

Stage Count =	959	700	259
	Overall Accuracy	Accurate NO Prediction	Accurate YES Prediction
Null Hypothesis	73.0%	100.0%	0.0%
$x_{CD} = 1500$ ft, Standard	79.7%	91.4%	47.9%
$x_{CD} = 2000$ ft, Standard	80.0%	79.1%	82.2%
$x_{CD} = 2500$ ft, Standard	64.8%	54.0%	93.8%
$x_{CD} = 1500$ ft, Modified	79.7%	91.6%	47.5%
$x_{CD} = 2000$ ft, Modified	83.8%	88.3%	71.8%
$x_{CD} = 2500$ ft, Modified	82.8%	85.4%	75.7%
Backpropagation NNET ¹	91.9%	96.4%	79.5%
Backpropagation NNET ²	92.9%	97.0%	81.9%

1 - backpropagation neural network with reservoir properties

2 - backpropagation neural network without reservoir properties

Table 13: Overall prediction accuracy comparison of all 5 models/methods investigated.

Conclusions

Artificial Intelligence proves to be superior at accurately predicting instances of frac hit occurrence when compared to current industry practices. Supervised learning techniques, such as the one used in this research, require ample data in order to train the models for accurately predicting blind datasets. To further build on this work it would be advantageous to increase the size of the dataset used and verify the distributions of the outcomes across each of the distance ranges.

The four most influential variables, in order of importance, were determined to be whether or not the child well stage and parent well stage fell within the frac plane, the distance between the child well and parent well, the existence of a producing well acting as a shield between the child well and parent well, and the number of days the parent well was producing prior to the offset child well completion. **Table 14** shows the degree of influence that each of these variables had in the model with respect to the next. For reference, OnPlane had the highest degree of influence which is assigned a value of 1 and used as the benchmark. All other variables are then compared to the benchmark to determine their relative degree of influence.

Predictor	% Degree of Influence
OnPlane	100%
MIN3D_Distance	96%
Shielded	38%
Delta_Days_HW Producing Before Hit	25%

Table 14: Degree of influence for the top four predictors in the final model.

The minimum stress lies in a N45E direction which is identified as the frac plane. Wells are drilled perpendicular to this plane to take full advantage of fracture geometry. This research shows that a frac hit occurrence is highly unlikely when a child well stage and parent well stage fall outside of this plane, which was defined as a +/- 10 degree window for this research.

Distance between wells is the 2nd most influential predictor. This can be determined with or without the use of neural networks. **Figure 11** shows that the percentage of frac hits decreases as the distance between wells increases. This is a simple correlation to pull from the dataset and an excellent verification that the neural network models are capable of learning like humans.

The predictors of Shielded and Delta_Days_HW Producing Before Hit are tied closely together. Delta_Days_HW Producing Before Hit is used as an analog for pressure depletion within the SRV of the parent well. Longer production days translates into greater pressure depletion, making the parent well more susceptible to a frac hit. A producing well acting as a Shield functions in the same manner. The pressure depletion from the well acting as a shield takes on most, if not all, of the child well frac energy thus protecting the potential parent well from being impacted.

Several ideas come to mind that could be completed to expand on this research for future research projects. The use of DTS and/or DAS could be used in the parent wells to remove the uncertainty of the assumption that the wells impacted along the frac plane. Although considered to be a good assumption, assumptions are never desired when performing data analytics. The concept of shielded wells could be further enhanced. For instance, in this research we only designated a well to be shield as a YES or a NO. Now knowing that pressure depletion has influence on the predictions, it may have been advantageous to record the well that was acting as a shield and include the pressure depletion analog for the shielded well as an input to the models. Finally, studying the completion parameters from the child well, shown in **Table 7**, should be investigated. These variables were removed from our dataset due to these being unavailable at the time a frac was initiated. However, inclusion of these variables could likely enhance the accuracy of the models. The operator could then perform sensitivity studies with known completion types to determine which completion best suits each individual stage within a child well, thus driving operational efficiencies to mitigate frac hits.

Definition of Terms and Acronyms

Child Well – The child well is the new drilled well where the stimulation, or frac, of this well has an opportunity to interfere with an offset producing well (Parent well). The designation of FW was used in the dataset to denote data variables stemming from the child well.

On Plane – On plane references the preferential direction in which a fracture will propagate in relation to the maximum and minimum stresses. The studied field frac plane is approximately N45E. The upper and lower limits for this research are N55E and N35E. If angle between the child well stage and parent well stage falls within this plane then the stages are consider “On Plane” represented by the variable 1. All others that fall outside of this plane are consider not to be on plane and designated as a 0.

Parent Well – The parent well is an existing producing well that has an opportunity to be interfered with by an offset well(s) completion (Child well). Multiple parent wells may exist for a single child well. The designation of HW was used in the dataset to denote data variables stemming from the parent well.

SRV – Stimulated Reservoir Volume

DAS – Distributed Acoustic Sensing

DTS – Distributed Temperature Sensing

x_{CD} – User defined “cutoff distance” that is believed to be the maximum reasonable distance that a frac hit can occur based on an individual’s experience in the production field .

RMSE – Root Mean Square Error. Used to determine the accuracy of the neural network predictions compared to the known outcomes within each epoch. The backpropagation algorithms objective is to reduce the RMSE with each additional epoch.

Epoch – The process of back propagating through a neural network changing weights, then using these new weights to forward propagate through the network to calculate a new output is called an Epoch. Neural network models run numerous (hundreds, thousands, or can be millions) epochs with the objective of reducing the RMSE to zero.

Input Layer – Refers to the first layer in the backpropagation network. The inputs, also referred to as predictors, are determined as part of the dataset collection and preparation.

Hidden Layer – Layer in a backpropagation neural network that lies between the input layer and output layer.

Activation Function – Activation Functions are used to convert an input signal to an output signal. The converted output is then used as the input for the proceeding layer in the model. Without the use of an activation function against the computed output signal the neural network model would act as a simple linear regression.

Model Bias – Bias allows the activation function to be shifted in order to better fit the data. Without bias, the activation function will always pass through origin.

Weights – Weights are assigned to each node to node relationship in the network. These values constrain how input data relates to output data. The weights are represented by the individual lines between nodes in **Figure 17**.

Training Set – The dataset is divided into 3 sections for processing using neural networks. The training set is used to train the network to reduce the RMSE. For this research the training set was user defined to be 80% of the total data points (767 rows).

Calibration Set – The weights and bias from the training set are used to calculate the RMSE for the calibration set. No training is completed against the calibration set as this dataset is meant to be a blind test of the backpropagation model to determine its effectiveness. For this research the calibration set was user defined to be 10% of the data points (96 rows).

Verification Set – The weights and bias from the training set are used to calculate the RMSE for the verification set. No training is completed against the verification set as this dataset is meant to be a **double** blind test of the backpropagation model to determine its effectiveness. For this research the verification set was user defined to be 10% of the data points (96 rows).

Cited Literature

Gidley, J.L., Holditch, S.A., Nierode, D.E. et al. 1989. An Overview of Hydraulic Fracturing. In Recent Advances in Hydraulic Fracturing, 12. Chap. 1, 1-38. Richardson, Texas: Monograph Series, SPE

Jacobs, T. (2017, April 1). Oil and Gas Producers Find Frac Hits in Shale Wells a Major Challenge. Society of Petroleum Engineers. doi:10.2118/0417-0029-JPT

<https://www.eia.gov/todayinenergy/detail.php?id=26112>

Lawal, H., Jackson, G., Abolo, N., & Flores, C. (2013, June 10). A Novel Approach To Modeling and Forecasting Frac Hits In Shale Gas Wells. Society of Petroleum Engineers. doi:10.2118/164898-MS

Paryani, M., Smaoui, R., Poludasu, S., Attia, B., Umholtz, N., Ahmed, I., & Ouenes, A. (2017, February 15). Adaptive Fracturing to Avoid Frac Hits and Interference: A Wolfcamp Shale Case Study. Society of Petroleum Engineers. doi:10.2118/185044-MS

Ajani, A. A., & Kelkar, M. G. (2012, January 1). Interference Study in Shale Plays. Society of Petroleum Engineers. doi:10.2118/151045-MS

Esquivel, R., & Blasingame, T. A. (2017, July 24). Optimizing the Development of the Haynesville Shale - Lessons Learned from Well-to-Well Hydraulic Fracture Interference. Unconventional Resources Technology Conference

Mukherjee, H., Poe, B. D., Heidt, J. H., Watson, T. B., & Barree, R. D. (2000, August 1). Effect of Pressure Depletion on Fracture-Geometry Evolution and Production Performance. Society of Petroleum Engineers. doi:10.2118/65064-PA

Miller, M. A., Jenkins, C. D., & Rai, R. R. (2010, January 1). Applying Innovative Production Modeling Techniques to Quantify Fracture Characteristics, Reservoir Properties, and Well Performance in Shale Gas Reservoirs. Society of Petroleum Engineers. doi:10.2118/139097-MS

King, G. E., Rainbolt, M. F., & Swanson, C. (2017, October 9). Frac Hit Induced Production Losses: Evaluating Root Causes, Damage Location, Possible Prevention Methods and Success of Remedial Treatments. Society of Petroleum Engineers. doi:10.2118/187192-MS

Mohaghegh, S. D. (2017, August 7). Mapping the Natural Fracture Network in Utica Shale Using Artificial Intelligence (AI). Unconventional Resources Technology Conference.

Alakbari, F. S., Elkatatny, S., & Baarimah, S. O. (2016, November 30). Prediction of Bubble Point Pressure Using Artificial Intelligence AI Techniques. Society of Petroleum Engineers. doi:10.2118/184208-MS

Okpo, E. E., Dosunmu, A., & Odagme, B. S. (2016, August 2). Artificial Neural Network Model for Predicting Wellbore Instability. Society of Petroleum Engineers. doi:10.2118/184371-MS

<https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>

<https://ihsmarkit.com/btp/fekete.html>

Appendix A – Definitions of Dataset Columns

FW – Frac Well. This designation is used across all of the variables that reference data points originating from the completed well. Better known in industry and referenced in this work as the Child Well.

HW – Hit Well. This designation is used across all of the variables that reference data points originating from the producing well. Better known in industry and referenced in this work as the Parent Well.

1. **Stage Frac** – ID of the frac stage. The ID includes an acronym for the well, well number, and stage number. I.e ABC-2-9
2. **Stage Hit** – ID of the stage in the producing well. Designation is same as Stage Frac
3. **Location** – Represented as an integer 1-6 based on geographical area of the stage. See the figure below are visual groupings of the actual dataset.

Figure 26: Classification of Stage Location

4. **FW_UTM X** – Universal Transverse Mercator coordinate system X coordinate. This coordinate represents the midpoint location of the stage.

5. FW_UTM Y – Universal Transverse Mercator coordinate system Y coordinate. This coordinate represents the midpoint location of the stage.
6. FW_Depth – Adjusted true vertical depth (ft) of the frac well to account for sea level
7. FW_MD – Total drilled depth (ft) of the frac well which includes the vertical section and lateral. Also known as Measured Depth
8. FW_KB – Datum used for measured depth (ft) and true vertical depth measurements. Kelly Bushing height includes sea level and rig floor height to which the measurements were taken.
9. FW_TVD – True Vertical Depth (ft). This variable represents the maximum vertical depth the wellbore reached at any point along the entire length of the wellbore.
10. FW_TVDSS – True Vertical Depth Sub Sea (ft). This variable is the same as FW_Depth but is calculated as FW_KB – FW_TVD as a double check. This depth is calculating using ground elevation as a datum. This provides a much clearer picture of vertical spacing between well laterals.
11. FW_Zone – The zone is the strata to which the stage was completed within represented as an integer 1-6.
12. FW_Fault – Indicates of a fault line intersects the stage. The value is represented as a 0 for no fault or 1 when a fault exists.
13. FW_Percent_of_LL – Ranges from 0% - 100% with 100% representing the first stage, pumped into the toe of the well, and 0% representing the last stage, pumped into the heel of the well.
14. FW Frac Gradient – Formation fracturing pressure in units of psi/ft.
15. FW Max Frac Pressure – Maximum treating pressure measured at surface during the treatment of the frac stage in units of psig.
16. FW Avg Frac Pressure – Average treating pressure measured at surface over the course of the entire stage in units of psig.
17. FW BrkD Frac Pressure – The pressure at which the formation broke down and began taking fluid in units of psig.
18. FW Max Frac Rate BPM – Maximum fluid pump rate during the frac stage in units of bbl/min.
19. FW Avg Frac Rate BPM – Average pump rate recorded over the course of the entire stage in units of bbl/min.
20. FW BrkD Frac Rate BPM – Fluid rate being pumped at the time the formation broke down in units of bbl/min.

21. FW_Total_Sand lbs – Total quantity of proppant pumped during the entire course of the stage in units of lbs.
22. FW_%_Fine_Sand – Percentage of total sand pumped that is considered 100 mesh. The remainder is 30/50.
23. FW_CLNVol bbls – Total fluid volume pumped during the stage in units of barrels.
24. FW_SLRYVol bbls – Total slurry volume pumped during the stage in units of barrels. Slurry is the mixture of water/sand as one fluid.
25. FW_Max Prop Conc PPG – Max proppant concentration reached during the course of the stage in units of lbs/gal.
26. FW_NetH_TOTAL – Net Pay of the target interval at the physical location of the stage within the formation in units of ft. This value is subject to change over the length of the entire lateral because horizontal wells can extend thousands of feet within the zone
27. FW_Total_TOC – Total Organic Carbons of the target interval at the physical location of the stage in units of %. Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
28. FW_Total_DEN – Formation density of the target interval at the physical location of the stage in units of g/cm³.
29. FW_Total_PHIE – Formation porosity of the target interval at the physical location of the stage in units of %. Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
30. FW_Total_PERM – Formation permeability of the target interval at the physical location of the stage in units of micro darcy (uD). Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
31. FW_Total_SW – Formation water saturation of the target interval at the physical location of the stages in units of %. Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
32. FW Ro Marcellus – Vitrinite reflectance of the target formation interval at the physical location of the stage in units of %. Ro is a measure of thermal maturity which is an indicator of whether the rock will produce oil, oil and gas, or gas only. Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
33. FW_BTU_Marcellus – Higher Heating Value of the target formation interval at the physical location of the stages in units of BTU (British Thermal Units). Measured via gas analysis at the wellhead and correlated horizontally, across the field, using sophisticated reservoir simulation software.

34. FW_PL_Marcellus – Langmuir pressure of the target formation interval at the physical location of the stage in units of psig. The pressure at which $\frac{1}{2}$ of Langmuir volume can be adsorbed. This value is derived from lab measurements.
35. FW_VL_Marcellus – Langmuir volume of the target formation interval at the physical location of the stage in units of scf/ton normalized by the % TOC of the sample. Langmuir volume is the maximum volume of gas that can be adsorbed to a coal or shale at infinite pressure. This value is derived from lab measurements.

Figure 27: Langmuir Volume (www.fekete.com)

36. FW_Pg_Marcellus – Pressure gradient of the target formation interval at the physical location of the stage in units of psi/ft. This value is calculated by dividing the true vertical depth of the well by the reservoir pressure.
37. FW_Pressure_Marcellus – Reservoir pressure of the target formation interval at the physical location of the stage in units of psig. This value is determined with DFIT testing and analysis and correlated horizontally using sophisticated reservoir simulation software.
38. FW_Temp_max_Marcellus – Maximum reservoir temperature of the target formation interval at the physical location of the stage in units of Fahrenheit. This value is measured using logs and correlated horizontally using sophisticated reservoir simulation software.
39. FW_Total_GIP – Gas In Place for the target formation interval at the physical location of the stage in units of BCF/mile². GIP is calculated using a series of log derived and lab derived variables and consists of adsorbed gas and free gas in formation.
40. Delta_Days_HW Producing Before Hit – The total number of days the parent well was on production prior to being hit by the child well.
41. MIN3D_Distance – The straight line distance measured between the child well stage being frac'd and the closet stage in the parent well. This distance is measured from midpoint stage to midpoint stage in units of feet.

42. TVD_Difference – The measured vertical distance between the child well stage being frac'd and the closet stage in the parent well. This distance is measured from midpoint stage to midpoint stage in units of feet.
43. OnPlane – Indicates if the parent well stage is located in the frac plane of the child well stage. This is represented by a 0 for “not on plane” or 1 for “on plane”. On plane references the preferential direction in which a fracture will grow in relation to the maximum and minimum stresses. If the child well fracture interferes with the parent well within this plane then the frac is consider “On Plane”. Pictorial representation of how this field is determined is shown below:

On Plane Illustration

Figure 28: Illustration showing how the "OnPlane" data point is derived.

44. Shielded – Indicates if the parent well stage and child well stage are direct offsets or if a producing wellbore exists in between the two wells. A child well and parent well with no producing well between them is defined as a 0 for “not shielded”. A child well and parent well with a producing well between them is defined as a 1 for “shielded”. The illustration below explains how this is determined.

Shielded Illustration

Figure 29: Illustration explaining how the “Shielded” data point is derived

45. **Effects** – This is the lone output field that we are modeling all other variables against in order to accurately predict the effect. The effect is designated as a 0 for “No Hit” or 1 for “Hit”. Effects was determined by manually reviewing the individual stages on the child well plotted against the parent well production profile. A frac hit is recognized by sharp deviation in casing pressure that corresponds with the child well frac stage

Figure 30: Definition for the output of Effects.

46. **CleanBBLsPumpedAtHit** – Total volume of water that had been pumped, in units of bbls, in the child well at the exact time the parent well showed a pressure response, indicating a hit. This data point was manually interpolated by looking at the child well treatment data.
47. **HW_UTM X** – Universal Transverse Mercator coordinate system X coordinate. This coordinate represents the midpoint location of the stage.
48. **HW_UTM Y** – Universal Transverse Mercator coordinate system Y coordinate. This coordinate represents the midpoint location of the stage.
49. **HW_Zone** – The zone is the strata to which the stage was completed within represented as an integer 1-6.
50. **HW_Fault** – Indicates if a fault line intersects the stage. The value is represented as a 0 for no fault or 1 when a fault exists.

51. HW_KB – Datum used for measured depth (ft) and true vertical depth measurements. Kelly Bushing height includes sea level and rig floor height to which the measurements were taken.
52. HW_TVD – True Vertical Depth (ft). This variable represents the maximum vertical depth the wellbore reached at any point along the entire length of the wellbore.
53. HW_TVDSS - True Vertical Depth Sub Sea (ft). This depth is calculating using ground elevation as a datum. This provides a much clearer picture of vertical spacing between well laterals.
54. HW_Percent_of_LL – Ranges from 0% - 100% with 100% representing the first stage, pumped into the toe of the well, and 0% representing the last stage, pumped into the heel of the well.
55. HW Frac Gradient – Formation fracturing pressure in units of psi/ft.
56. HW Max Frac Pressure – Maximum treating pressure measured at surface during the treatment of the frac stage in units of psig.
57. HW Avg Frac Pressure – Average treating pressure measured at surface over the course of the entire stage in units of psig.
58. HW BrkD Frac Pressure – The pressure at which the formation broke down and began taking fluid in units of psig.
59. HW Max Frac Rate BPM – Maximum fluid pump rate during the frac stage in units of bbl/min.
60. HW Avg Frac Rate BPM – Average pump rate recorded over the course of the entire stage in units of bbl/min.
61. HW BrkD Frac Rate BPM – Fluid rate being pumped at the time the formation broke down in units of bbl/min.
62. HW_Total_Sand lbs – Total quantity of proppant pumped during the entire course of the stage in units of lbs.
63. HW_%_Fine_Sand – Percentage of total sand pumped that is considered 100 mesh. The remainder is 30/50.
64. HW_CLNVol bbls – Total fluid volume pumped during the stage in units of barrels.
65. HW_SLRYVol bbls – Total slurry volume pumped during the stage in units of barrels. Slurry is the mixture of water/sand as one fluid.
66. HW_Max Prop Conc PPG – Max proppant concentration reached during the course of the stage in units of lbs/gal.
67. HW_NetH_TOTAL – Net Pay of the target interval at the physical location of the stage within the formation in units of ft. This value is subject to change over the length of the entire lateral because horizontal wells can extend thousands of feet within the zone

68. HW_Total_TOC – Total Organic Carbons of the target interval at the physical location of the stage in units of %. Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
69. HW_Total_DEN – Formation density of the target interval at the physical location of the stage in units of g/cm³.
70. HW_Total_PHIE – Formation porosity of the target interval at the physical location of the stage in units of %. Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
71. HW_Total_PERM – Formation permeability of the target interval at the physical location of the stage in units of micro darcy (uD). Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
72. HW_Total_SW – Formation water saturation of the target interval at the physical location of the stages in units of %. Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
73. HW Ro Marcellus – Vitrinite reflectance of the target formation interval at the physical location of the stage in units of %. Ro is a measure of thermal maturity which is an indicator of whether the rock will produce oil, oil and gas, or gas only. Measured on a vertical log and correlated horizontally with sophisticated reservoir simulation software.
74. HW_BTU_Marcellus – Higher Heating Value of the target formation interval at the physical location of the stages in units of BTU (British Thermal Units). Measured via gas analysis at the wellhead and correlated horizontally, across the field, using sophisticated reservoir simulation software.
75. HW_PL_Marcellus – Langmuir pressure of the target formation interval at the physical location of the stage in units of psig. The pressure at which ½ of Langmuir volume can be adsorbed. This value is derived from lab measurements.
76. HW_VL_Marcellus – Langmuir volume of the target formation interval at the physical location of the stage in units of scf/ton normalized by the % TOC of the sample. Langmuir volume is the maximum volume of gas that can be adsorbed to a coal or shale at infinite pressure. This value is derived from lab measurements. See Figure **Figure 27: Langmuir Volume (www.fekete.com)** for a visual explanation.
77. HW_Pg_Marcellus – Pressure gradient of the target formation interval at the physical location of the stage in units of psi/ft. This value is calculated by dividing the true vertical depth of the well by the reservoir pressure.
78. HW_Pressure_Marcellus – Reservoir pressure of the target formation interval at the physical location of the stage in units of psig. This value is determined with DFIT testing and analysis and correlated horizontally using sophisticated reservoir simulation software.

79. HW_Temp_max_Marcellus – Maximum reservoir temperature of the target formation interval at the physical location of the stage in units of Fahrenheit. This value is measured using logs and correlated horizontally using sophisticated reservoir simulation software.
80. HW_Total_GIP – Gas In Place for the target formation interval at the physical location of the stage in units of BCF/mile². GIP is calculated using a series of log derived and lab derived variables and consists of adsorbed gas and free gas in formation.

Appendix B – Backpropagation Neural Network Models

Figure 31 shows the final model that was used. The blue highlights represent the predictors that were used in the model. **Green** lettering indicates the unique identifiers for each record and **red** lettering represents the output,

Stage Frac	FW_Total_DEN	HW Frac Gradient
Stage Hit	FW_Total_PHIE	HW Max Frac Pressure
Location	FW_Total_PERM	HW Avg Frac Pressure
FW_UTM X	FW_Total_SW	HW BrkD Frac Pressure
FW_UTM Y	FW_Ro_Marcellus	HW Max Frac Rate BPM
FW_Depth	FW_BTU_Marcellus	HW Avg Frac Rate BPM
FW_MD	FW_PL_Marcellus	HW BrkD Frac Rate BPM
FW_KB	FW_VL_Marcellus	HW_Total_Sand lbs
FW_TVD	FW_Pg_Marcellus	HW_%_Fine_Sand
FW_TVDSS	FW_Pressure_Marcellus	HW_CLNVol bbls
FW_Zone	FW_Temp_max_Marcellus	HW_SLRYVol bbls
FW_Fault	FW_Total_GIP	HW_Max Prop Conc PPG
FW_Percent_of_LL	Delta_Days_HW Producing Before Hit	HW_Neth_TOTAL
FW Frac Gradient	MIN3D_Distance	HW_Total_TOC
FW Max Frac Pressure	TVD_Difference	HW_Total_DEN
FW Avg Frac Pressure	OnPlane	HW_Total_PHIE
FW BrkD Frac Pressure	Shielded	HW_Total_PERM
FW Max Frac Rate BPM	Effects	HW_Total_SW
FW Avg Frac Rate BPM	CleanBBLsPumpedAtHit	HW_Ro_Marcellus
FW BrkD Frac Rate BPM	HW_UTM X	HW_BTU_Marcellus
FW_Total_Sand lbs	HW_UTM Y	HW_PL_Marcellus
FW_%_Fine_Sand	HW_Zone	HW_VL_Marcellus
FW_CLNVol bbls	HW_Fault	HW_Pg_Marcellus
FW_SLRYVol bbls	HW_KB	HW_Pressure_Marcellus
FW_Max Prop Conc PPG	HW_TVD	HW_Temp_max_Marcellus
FW_Neth_TOTAL	HW_TVDSS	HW_Total_GIP
FW_Total_TOC	HW_Percent_of_LL	

Figure 31: Complete list of columns collected as part of the data mining process. The predictors used in the final model are highlighted in blue.

Figure 32 shows the variation of the final model that contained reservoir properties for both the parent well and child well. This model predicted with more confidence, but no increase in accuracy. Reservoir properties are generally harder to obtain throughout the data mining process. This research showed that for the studied reservoir that having these properties added little value to the overall work.

Stage Frac	FW_Total_DEN	HW Frac Gradient
Stage Hit	FW_Total_PHIE	HW Max Frac Pressure
Location	FW_Total_PERM	HW Avg Frac Pressure
FW_UTM X	FW_Total_SW	HW BrkD Frac Pressure
FW_UTM Y	FW Ro Marcellus	HW Max Frac Rate BPM
FW_Depth	FW_BTU_Marcellus	HW Avg Frac Rate BPM
FW_MD	FW_PL_Marcellus	HW BrkD Frac Rate BPM
FW_KB	FW_VL_Marcellus	HW_Total_Sand lbs
FW_TVD	FW_Pg_Marcellus	HW_%_Fine_Sand
FW_TVDSS	FW_Pressure_Marcellus	HW_CLNVol bbls
FW_Zone	FW_Temp_max_Marcellus	HW_SLRYVol bbls
FW_Fault	FW_Total_GIP	HW_Max Prop Conc PPG
FW_Percent_of_LL	Delta_Days_HW Producing Before Hit	HW_NetH_TOTAL
FW Frac Gradient	MIN3D_Distance	HW_Total_TOC
FW Max Frac Pressure	TVD_Difference	HW_Total_DEN
FW Avg Frac Pressure	OnPlane	HW_Total_PHIE
FW BrkD Frac Pressure	Shielded	HW_Total_PERM
FW Max Frac Rate BPM	Effects	HW_Total_SW
FW Avg Frac Rate BPM	CleanBBLsPumpedAtHit	HW Ro Marcellus
FW BrkD Frac Rate BPM	HW_UTM X	HW_BTU_Marcellus
FW_Total_Sand lbs	HW_UTM Y	HW_PL_Marcellus
FW_%_Fine_Sand	HW_Zone	HW_VL_Marcellus
FW_CLNVol bbls	HW_Fault	HW_Pg_Marcellus
FW_SLRYVol bbls	HW_KB	HW_Pressure_Marcellus
FW_Max Prop Conc PPG	HW_TVD	HW_Temp_max_Marcellus
FW_NetH_TOTAL	HW_TVDSS	HW_Total_GIP
FW_Total_TOC	HW_Percent_of_LL	

Figure 32: Variation of the final model used that contains reservoir properties. This model predicted frac hits with more confidence but ultimately resulted in no additional accuracy.

Figure 33 shows the breakout of training data vs the calibration and verification data for the model with reservoir properties.

Figure 33: Training data and Calibration/Verification dataset accuracies for the backpropagation neural network model with reservoir properties.

Figure 34 shows the breakout of training data vs the calibration and verification data for the model that does not include reservoir properties.

Figure 34: Training data and Calibration/Verification dataset accuracies for the backpropagation neural network model that does not contain reservoir properties.