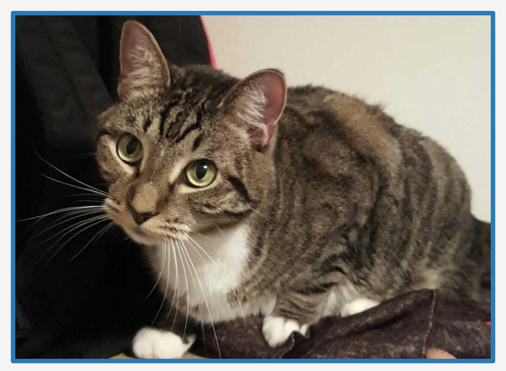
An Evaluation of Machine Learning Algorithms for Classification of Shelter Animal Outcomes

Mikayla Timm and Eman El-Sheikh, Ph.D.
Department of Computer Science
University of West Florida

University of West Florida
Computer Science

Introduction

Background



- Every year, **animal shelters** across the U.S. give approximately **7.6 million animals** a chance at finding a forever home and starting a new life [1].
- **Certain attributes** of the animals recorded at shelter intake may **affect their outcomes**.
- Shelters can **focus their efforts** on animals who are less likely to be adopted if they know the animal is at risk of having a **negative outcome**.

Goals

- Develop a system utilizing **machine learning** algorithms to **predict the outcomes** of animals brought to shelters based on recorded attributes.
- Determine **which attributes** carry the **most significance** in classifying outcomes.

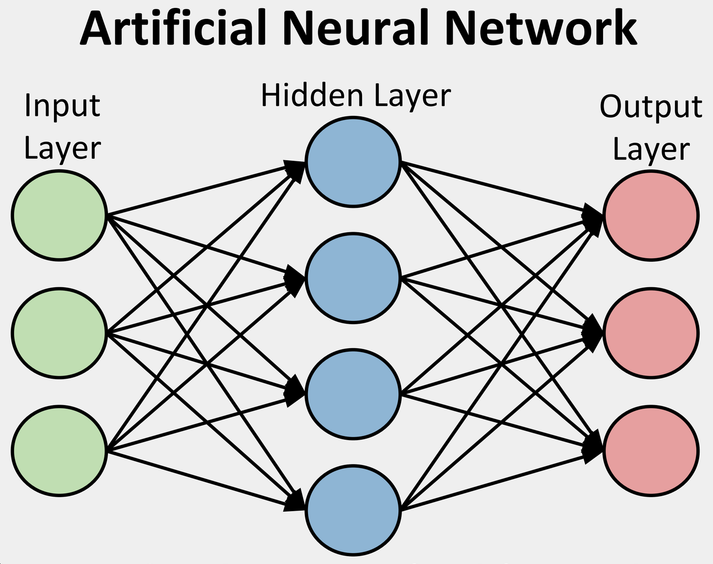
Approach

Data

- Data taken from Austin Animal Center from October, 2013, to March, 2016.
- **Attributes:** Name, Animal Type, Sex, Intactness, Age, Breed, Color, Date/Time
- **Possible Outcomes:** Adoption, Return to Owner, Transfer, Euthanasia, Died
- **Number of Animals in Training Set:** 26729
- **Number of Animals in Testing Set:** 11456

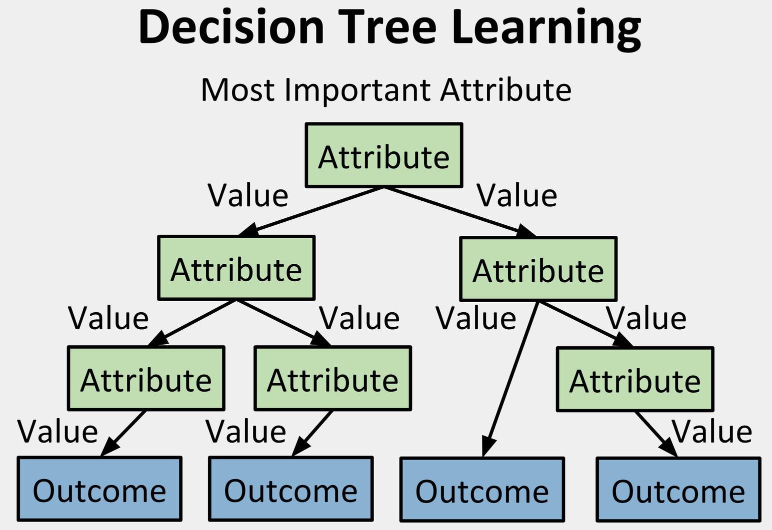
Classification

Artificial Neural Network



- **Weighted neurons** feed forward signals through hidden layers when weights surpass a **threshold** in an **activation function** [2], [3]
- Output layer: **outcomes**
- **Parameterized**
- Relatively **slow**
- **Complex** structure

Decision Tree Learning

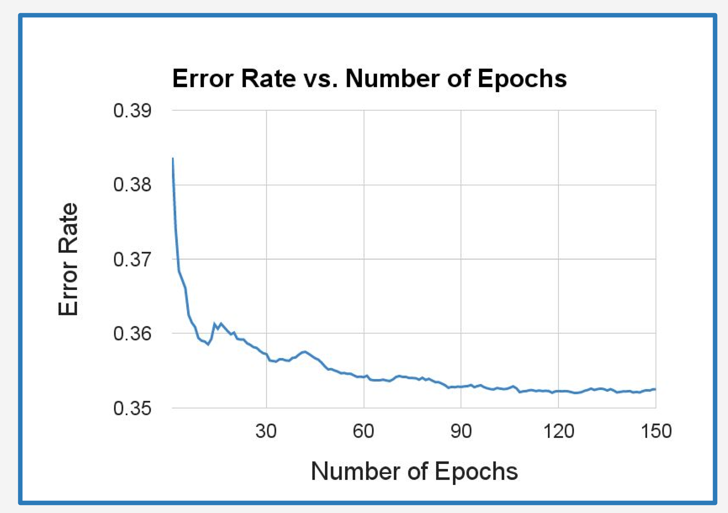


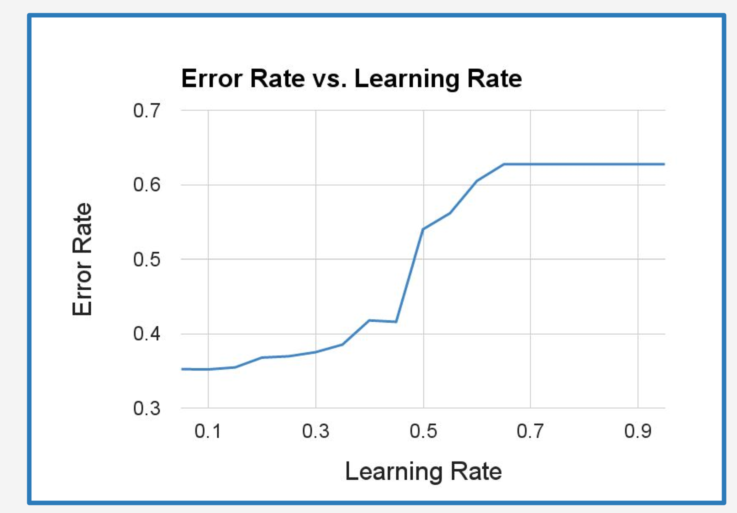
- **Splits instances** by **most important attribute** until all instances have **same classification** [2], [3]
- Most important attribute determined by **information gain** calculation
- **Not Parameterized**
- Relatively **fast**
- **Simple** structure

Results

Artificial Neural Network Parameterization

- By altering the **learning rate** and **number of epochs** for the neural network, we can observe changes in the **total error** for classification of the training set.

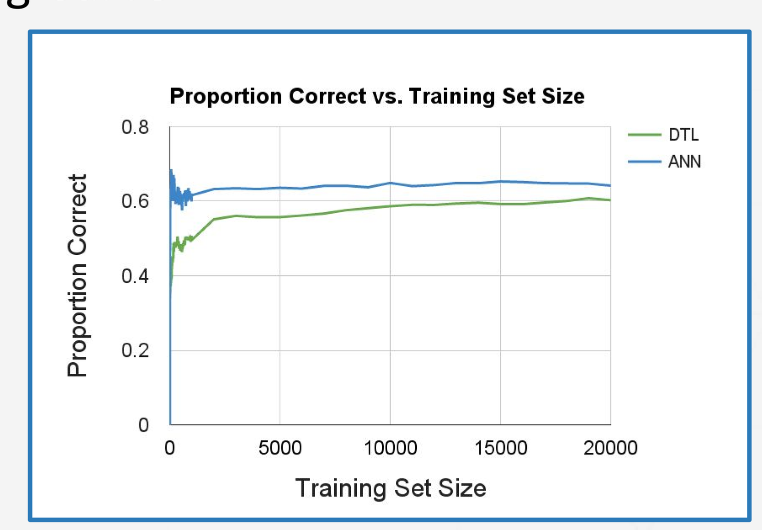




- Increasing the **number of epochs** leads to a **lower error rate** for the ANN.
- Increasing the **learning rate** to a point leads to a **higher error rate** for the ANN.

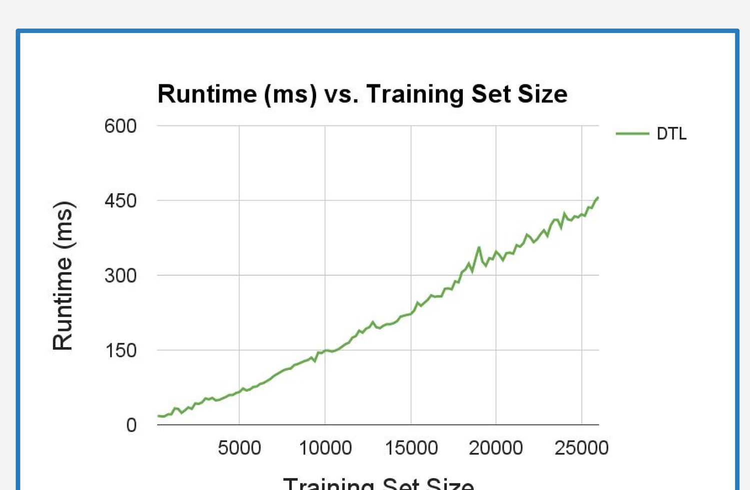
Artificial Neural Network (ANN) vs. Decision Tree Learning (DTL)

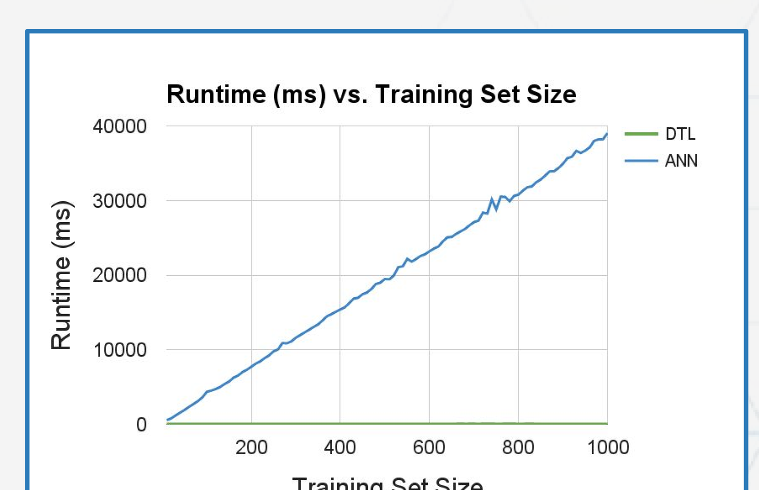
Learning Curve



- The **ANN outperforms** the DTL in terms of **accuracy**.
- The DTL's accuracy grows at a **slower initial rate**, but reaches a higher accuracy given **more examples**.

Runtime

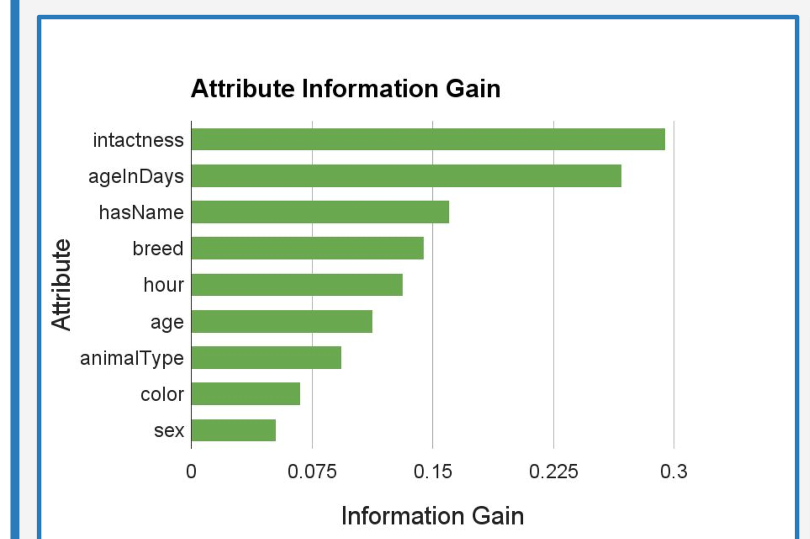


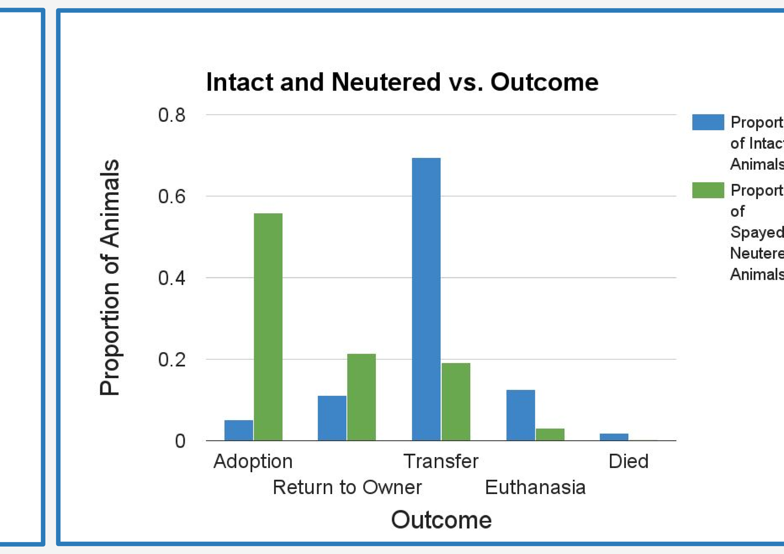


- Given many examples, the **DTL's runtime increases slightly**.
- However, the **ANN's runtime grows** at a much **faster** rate.

Results (cont.)

- By ranking the attributes by their **information gain**, we can see which **attributes** give the **most insight** into the final outcome.





- **Spayed/neutered** animals have **higher adoption** and **lower euthanization** rates.

Conclusion

Discussion

- The **ANN** classifies a **higher percentage of correct instances** than the DTL, but takes a **much longer time** to run (no free lunch).
- Given a greater number of **examples**, both algorithms **improve** in **accuracy**, though the runtime increases significantly for the ANN.
- Some **attributes** of the animals **do not matter much** (sex, color), while some make a **big difference** (intactness, ageInDays) in determining the outcomes.
- Some attributes cannot be controlled, but **adoption rates** may **increase** by **spaying/neutering** the shelter animals and giving each one a **name**.

Future Directions

Add decision tree **pruning** to combat overfitting issues	Implement **GUI** for viewing visualizations of attributes/outcomes	Test **other algorithms**, such as Naive Bayes, and compare results

References

1. **ASPCA**. Pet Statistics. http://www.aspca.org/animal-homelessness, 2016.
2. **M. Aly.** Survey on Multiclass Classification Methods. 2005.
3. **S. Russell and P. Norvig.** Artificial Intelligence A Modern Approach, 3rd Edition. 2010.

Acknowledgements

This project is inspired by the Kaggle competition "Shelter Animal Outcomes" on www.kaggle.com. The input data is provided by Austin Animal Center.
The Neural Network framework (Multilayer Perceptron) is provided by Weka.
Special thanks to **General Dynamics Information Technology** for sponsoring this research as part of UWF's Summer Undergraduate Research Program.