

ASCENDS: Advanced data SCiENce toolkit for Non-Data Scientists

Sangkeun Lee¹, Jian Peng¹, Andrew Williams², and Dongwon Shin¹

1 Oak Ridge National Laboratory 2 Cornell University

DOI: [10.21105/joss.01656](https://doi.org/10.21105/joss.01656)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Yuan Tang](#) ↗

Reviewers:

- [@zhampel](#)
- [@jrbourbeau](#)

Submitted: 24 July 2019

Published: 06 February 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Acknowledgements

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

The research was supported by the U. S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office, Propulsion Materials Program.

Summary

Recently, advances in machine learning and artificial intelligence have been playing more and more critical roles in a wide range of areas. For the last several years, industries have shown that how learning from data, identifying patterns and making decisions with minimal human intervention can be extremely useful to their business (e.g., image classification, recommending a product to a customer, finding friends in a social network, predicting customers actions, etc.). These success stories have been motivating scientists who study physics, chemistry, materials, medicine, and many other subjects, to explore a new pathway of utilizing machine learning techniques like regression and classification for their scientific activities. However, most existing machine learning tools, systems, and methodologies have been developed for programming experts but not for scientists (or any users) who have no or little knowledge of programming.

ASCENDS is a toolkit that is developed to assist scientists (or any persons) who want to use their data for machine learning tasks, more specifically, correlation analysis, regression, and classification. ASCENDS does not require programming skills. Instead, it provides a set of simple but powerful CLI (Command Line Interface) and GUI (Graphic User Interface) tools for non-data scientists to be able to intuitively perform advanced data analysis and machine learning techniques. ASCENDS has been implemented by wrapping around open-source software including Keras (Gulli & Pal, 2017), TensorFlow (Abadi et al., 2016), and scikit-learn (Pedregosa et al., 2011).

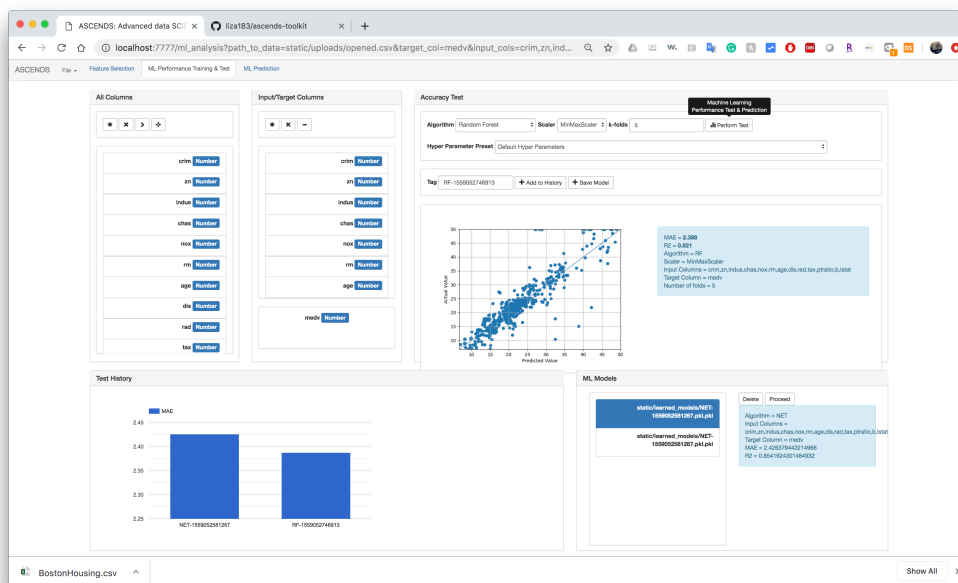


Figure 1: Using Ascends via its web-based graphic user interface

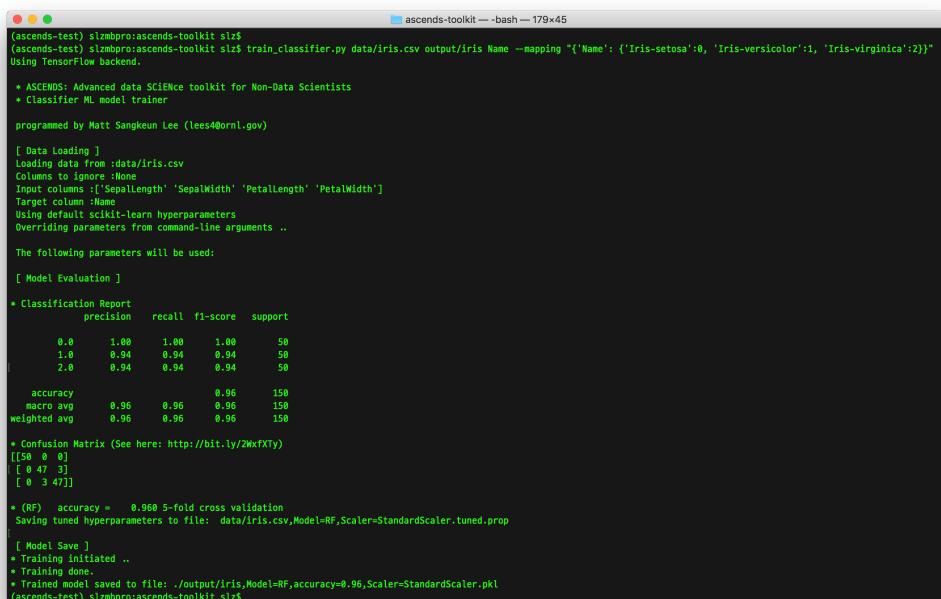


Figure 2: Using Ascends via its command-line interface

Users can perform three major tasks using ASCENDS as follows. First of all, users can easily perform correlation analysis (Ezekiel, 1930) using ASCENDS. ASCENDS can quantify the correlation between input variables (X) and an output variable (y) using various correlation coefficients including Pearson's correlation coefficient (Sedgwick, 2012) and maximal information coefficient (Reshef et al., 2011). Second, users can train, test, save and utilize classification models (Ren & Malik, 2003) without any programming efforts. For instance,

with ASCENDS, by executing a single command in a terminal, a user can train a model for predicting whether an email is a spam or not-spam. Last, similarly, users can train, test, save and use regression models (Darlington, 1990). For instance, ASCENDS can be used to train a model for predicting the value of a house based on square footage, number of bedrooms, number of cars that can be parked in its garages, and amount of storage using the provided graphic user interface in a web browser.

Earlier versions of ASCENDS have been used for scientific research such as Shin, Yamamoto, Brady, Lee, & Haynes (2019), Shin, Lee, Shyam, & Haynes (2017), and Wang et al. (2019).

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (pp. 265–283). Retrieved from <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>
- Darlington, R. B. (1990). *Regression and linear models*. McGraw-Hill New York.
- Ezekiel, M. (1930). *Methods of correlation analysis*.
- Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.
- Ren, X., & Malik, J. (2003). Learning a classification model for segmentation. In *Proceedings Ninth IEEE International Conference on Computer Vision* (p. 10). IEEE. doi:[10.1109/ICCV.2003.1238308](https://doi.org/10.1109/ICCV.2003.1238308)
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., et al. (2011). Detecting novel associations in large data sets. *Science*, *334*(6062), 1518–1524.
- Sedgwick, P. (2012). Pearson's correlation coefficient. *British Medical Journal*, *345*, e4483. doi:[10.1136/bmj.e4483](https://doi.org/10.1136/bmj.e4483)
- Shin, D., Lee, S., Shyam, A., & Haynes, J. A. (2017). Petascale supercomputing to accelerate the design of high-temperature alloys. *Science and Technology of Advanced Materials*, *18*(1), 828–838. doi:[10.1080/14686996.2017.1371559](https://doi.org/10.1080/14686996.2017.1371559)
- Shin, D., Yamamoto, Y., Brady, M. P., Lee, S., & Haynes, J. A. (2019). Modern data analytics approach to predict creep of high-temperature alloys. *Acta Materialia*, *168*, 321–330. doi:[10.1016/j.actamat.2019.02.017](https://doi.org/10.1016/j.actamat.2019.02.017)
- Wang, J., Nobakht, A. Y., Blanks, J. D., Shin, D., Lee, S., Shyam, A., Rezayat, H., et al. (2019). Machine learning for thermal transport analysis of aluminum alloys with precipitate morphology. *Advanced Theory and Simulations*. doi:[10.1002/adts.201800196](https://doi.org/10.1002/adts.201800196)