

Steven Boada, Ph.D

Contact Information	stevenboada@gmail.com (615) 200-0119	linkedin.com/in/theboada github.com/boada
Skills	<p>Machine Learning: Linear Models, Decision Trees, SVM, Clustering, Deep Learning, Feature Engineering</p> <p>Statistical Methods: Hypothesis testing and confidence intervals, error analysis, image analysis, Monte Carlo methods (e.g., emcee), maximum likelihood</p> <p>Software and Computing: Python (e.g. Scikit-learn, Numpy, Scipy, Pandas, Matplotlib, PyTorch), MySQL, ANSI C, Linux Command Line Environments, Microsoft Office Suite, GPGPU, and HPC applications</p> <p>Leadership: Demonstrable ability to tackle loosely defined problems, 5+ years organizing workflows from group planning sessions through implementation and delivery of final products, Eagle Scout.</p>	
Professional Experience	<p>Insight Data Science, New York, New York USA <i>Fellow</i> January, 2020 – Present</p> <ul style="list-style-type: none">• Addressed a shortage of NYC health inspectors which caused critical health violations to remain unaddressed for extended periods of time potentially harming the general public.• Trained a transparent model to prioritize NYC restaurants based on environmental variables and their past inspection histories.• Resulted in a XX% improved performance of NYC inspectors, leading to critical violations being discovered XX days earlier than previously expected.. <p>Dept. of Physics and Astronomy, Rutgers University, New Brunswick, New Jersey USA <i>Postdoctoral Research Associate</i> September, 2016 – 2020</p> <ul style="list-style-type: none">• Designed and built massive, parallelized, Python pipelines to process and analyze TBs of astronomical imaging; producing calibrated, standardized data catalogs and rigorous results.• Project managed and coordinated a team of 4, including both senior scientists and graduate students, to perform quality control tasks; deliver science products; and produce peer-reviewed publications. See: ArXiv:1808.06378• Contributed to widely used, open source, Python projects through bug fixes and feature additions: PHOTOMETRYPIPELINE, ASTLIB, and EASYGALAXY. <p>Texas A&M University, College Station, Texas USA <i>Ph.D Candidate</i> August, 2010 – 2016</p> <ul style="list-style-type: none">• Proved simulated results for an upcoming astronomical survey could be improved, by a factor of ~ 3, over in-house statistical methods by using Random Forest regression. Implemented these ML methods and produced improved results in a pilot survey of the real sky and under real-world conditions.• Collaborated with group members both in person, and through collaborative tools (e.g., GitHub, SVN).• Presented scientific results in high-impact, peer reviewed journals and at international conferences.	
Data Projects	<p>Using Imaging to Infer Galaxy Properties</p> <ul style="list-style-type: none">• Predicted galaxy chemical composition with $\sim 5\%$ error from pseudo-three color imaging, a result better than other current, similar efforts in the literature.• Leveraged Convolution Neural Networks, trained on GPUs, to analyze $\sim 150,000$ images from the Sloan Digital Sky Survey.• Project start to publication, 4 months. See ArXiv:1810.12913 <p>Predicting Tournament Performance in Warmachine</p> <ul style="list-style-type: none">• Created an Elo based model to forecast the results of an upcoming tournament and to identify potential future upsets.• Wrangled ~ 1800 tournament game results of a popular tabletop game using Python (e.g., Pandas).	
Education	<p>Texas A&M University, College Station, Texas</p> <ul style="list-style-type: none">• Ph.D, Physics (astronomy focus), August, 2016	<p>The University of Tennessee, Knoxville, Tennessee</p> <ul style="list-style-type: none">• M.S., Physics (astronomy focus), August, 2009• B.S., Physics, May, 2007