Using Recurrent Neural Networks in a Distributed Computing Environment for Predicting Time-Variant Data

Alani Seaman, Keely Mashburn, Tyler J. Wise

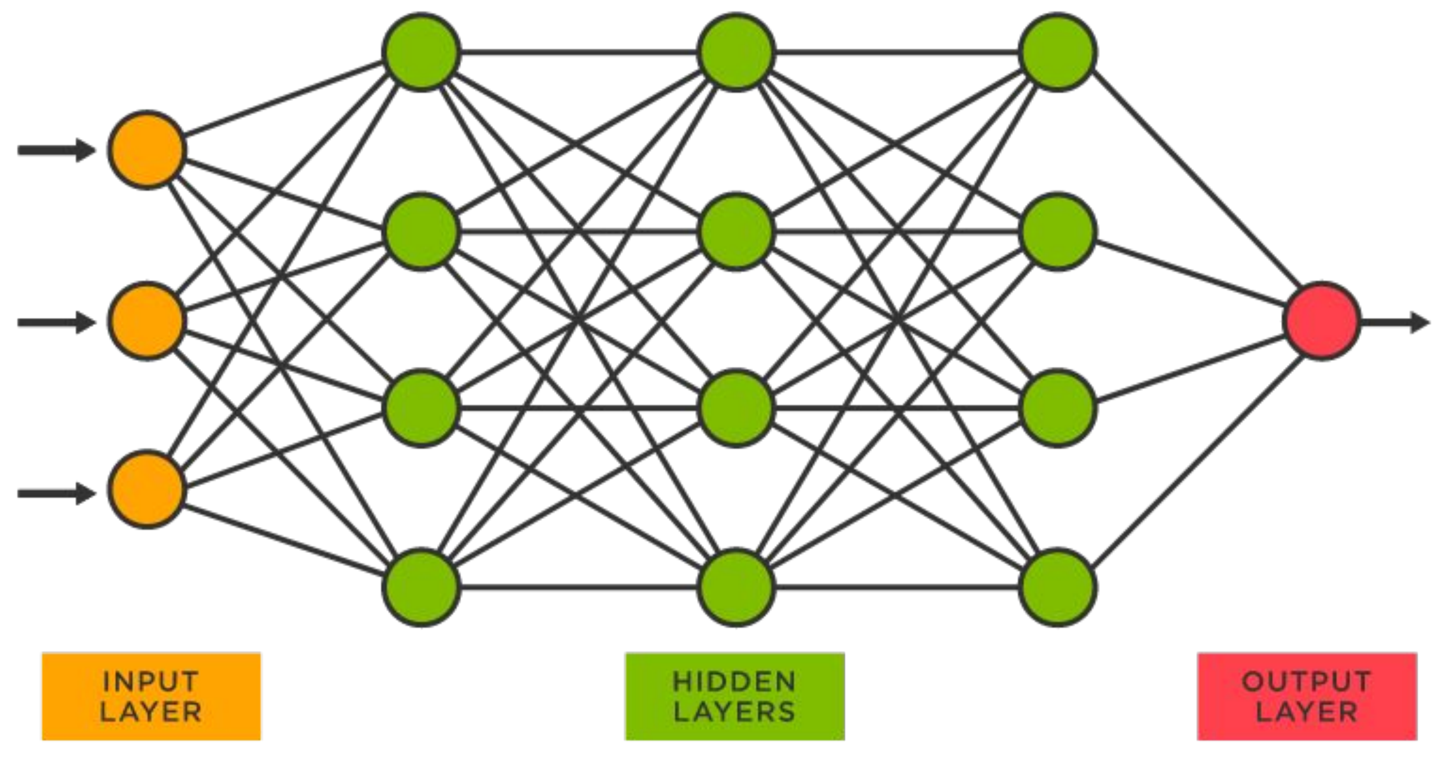
Abstract

By applying the concepts of machine learning, the aim is to create a program that utilizes neural networks to analyze the wait times at various Florida theme parks. These parks include SeaWorld, Busch Gardens, both Universal parks, and all four Walt Disney World parks. The project hinges on a distributed computing architecture that divides the work as assigned by a master, rather than strictly parallelizing the code. The technology used throughout this project is hosted on Amazon Web Services, utilizing their Relational Database Service and Sagemaker platforms. MySQL, Python, and Tensorflow are the core software technologies running on this infrastructure. Each of these programs plays a role in creating a complete solution towards creating a recurrent neural network that delivers a list of wait times synthesized for the following hour that users can benefit from in real time. For our analysis of the network's validity, we will create a statistical distribution for the error present in each ride's prediction. This will be performed on a testing data set, which is composed of twenty percent of the overall data chosen at random.

System Design

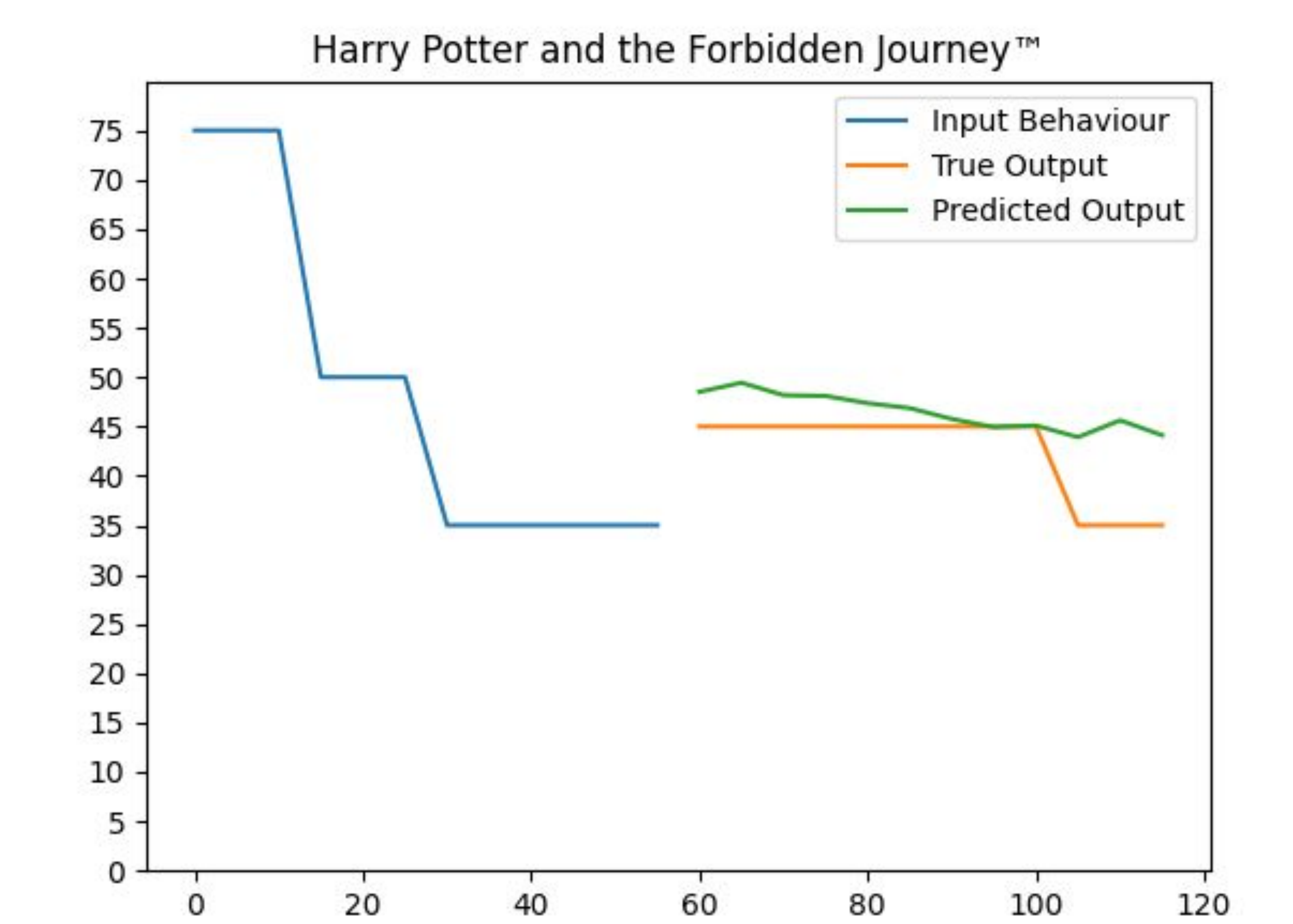
- This project utilized SageMaker, an Amazon cloud machine learning platform, to train the prediction model. This service runs using Amazon Web Services (AWS) and provides access to servers to run code and return the results.
- The python code developed to construct the models was stored in separate Jupyter notebooks.
- These notebooks connected directly to SageMaker and accessed the platform's machine-learning resources.

Neural Networks



System Accuracy

Ride Name	Accuracy
The Cat in the Hat	89%
Doctor Doom's Fearfall	72%
Despicable Me Minion Mayhem	55%
Hollywood Rip Ride Rockit	47%
Infinity Falls	94%
Kumba	96%
Toy Story Mania	61%
Star Tours - The Adventures Continue	70%
Expedition Everest	69%
DINOSAUR	59%
Journey Into Imagination with Figment	94%
Soarin' Around the World	58%
Jungle Cruise	60%
Haunted Mansion	74%



Analysis

- **Above:** This graph shows a sample output for one ride as part of the analysis. The line segment in blue shows the data for the rides previous hour of wait times, the yellow segment indicates what happened in the park in the following hour, and the green segment shows the algorithm's prediction.
- **Left:** This table shows a selection of accuracies for several rides at Universal Studios and Islands of Adventure parks. Accuracy is defined as the likelihood of a prediction set being correct within ten minutes.