

# Project Proposal

**Project Title:** Exploring Neural Networks algorithms for Handwritten Text Recognition using TensorFlow and Spark MLlib

**Team:**

USERNAME	FIRST NAME	LAST NAME
s5	Praveen	Surendran
pjagwani	Prashant	Jagwani
iklein	Irina	Klein
oamole	Oluwaseun	Amole

**Description:**

Handwritten Text Recognition (HTR), is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch-screens and other devices.

TensorFlow is a widely-used open-source platform for building and deploying Machine Learning models. MLlib is Spark's Machine Learning library. In this project we plan to implement and compare HTR algorithms using these tools to explore their differences and capabilities in practice.

**Goal:** Build an algorithm to successfully recognize handwritten digits and letters.

**Tools:** TensorFlow, Python, OpenCV, LMDB, Spark MLlib

**Plan:**

1. Research TensorFlow algorithms (Literature Review)
2. Build a neural network algorithm recognizing digits based on kaggle dataset <https://www.kaggle.com/competitions/digit-recognizer/code>
3. Research Spark MLlib algorithms (Literature Review)
4. Compare the implementation to Spark MLlib

**Literature:**

1. Article "Build a Handwritten Text Recognition System using TensorFlow"  
<https://towardsdatascience.com/build-a-handwritten-text-recognition-system-using-tensorflow-2326a3487cd5>
2. TensorFlow for Deep Learning - From Linear Regression to Reinforcement Learning  
<https://www.oreilly.com/library/view/tensorflow-for-deep/9781491980446/>
3. Handwritten Text Recognition in Historical Documents  
<https://repositum.tuwien.at/retrieve/10807>
4. Machine Learning with Apache Spark  
<https://www.oreilly.com/library/view/machine-learning-with/9781098106812/>
5. Deep Convolutional Generative Adversarial Network. TensorFlow  
<https://www.tensorflow.org/tutorials/generative/dcgan>
6. Generative Adversarial Networks. Cornell University  
<https://arxiv.org/abs/1406.2661>