



Deploying Machine Learning Application on Edge Devices, a Study of Challenges and Solutions

Students: Ernesto Vilches Martinez

Mentor: Luis Robaina, MITRE

Instructor/Faculty: Seyedmasoud Sadjadi, PhD, Florida International University

PROBLEM

Even though there has been a lot of innovation in the world of Artificial Intelligence, most of the current research efforts are focused on developing better algorithms and architectures while not much effort into understanding the engineering challenges of bringing AI systems to edge devices. My role in this project pertains to discovering different model types which are used for Edge devices and deploying them across a fleet of devices independent of application updates

CURRENT SYSTEM

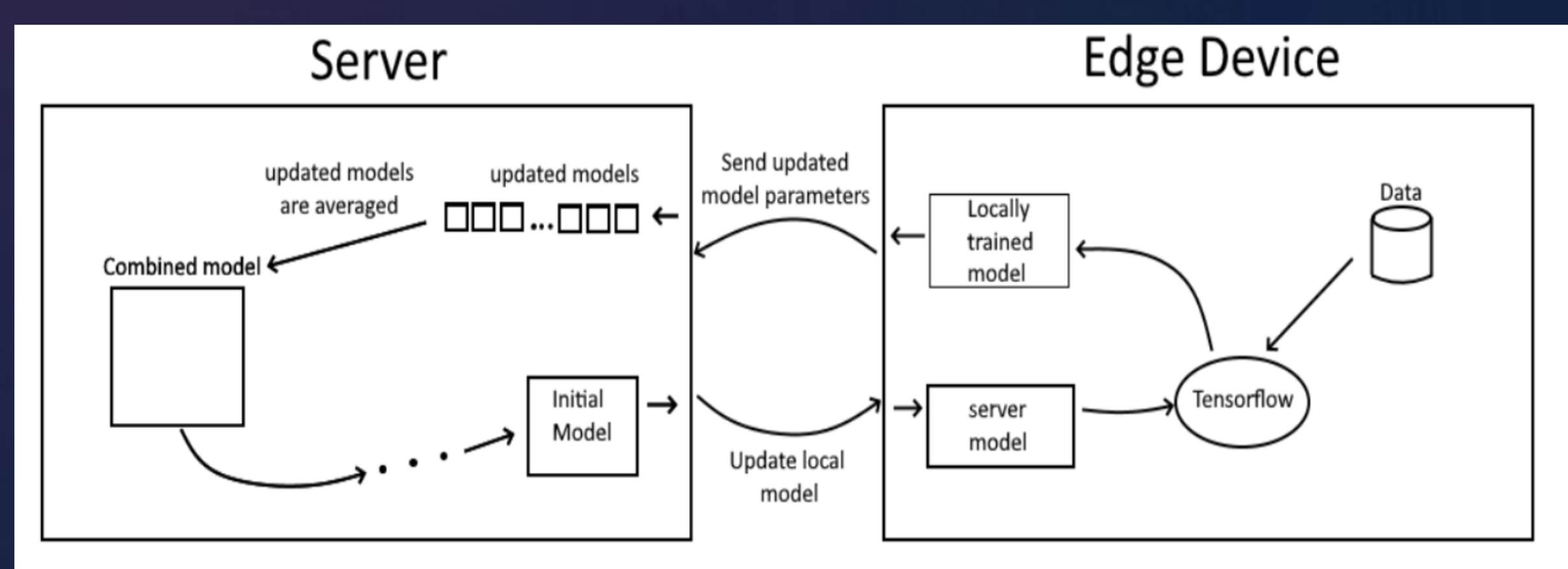
Edge devices are considered a fundamental part of distributed AI systems. Back in the day, technological devices were only able to manifest data by sending it to a cloud for further analysis. However, this has changed as now those very own IoT devices can perform their computations thanks to the major improvements we have made with software devices and ultimately resulting in edge computing. Edge computing has revolutionized how data is processed from billions of devices all over the world.

REQUIREMENTS

Contribute to the research paper by:

- Researching
 - The deployment of AI/ ML applications to edge devices.
 - How to monitor the AI/ ML applications once deployed on edge devices.
- Learning how to use LaTeX in VSCode to write all research.
- Uploading all research to GitHub.

SYSTEM DESIGN



This ML architecture model is a hybrid between fully centralized cloud computing and fully local edge training.

MITRE

LATEX



IMPLEMENTATION

- Team split into subgroups to research different topics regarding the project.
- We used GitHub to push all research/annotation and documents in relation to the research paper.
- VSCode was implemented to write the research for the paper with the help of LaTeX.
- Draw.io was used to create diagrams for the research paper.

VERIFICATION

Along with the help of the students in the Deployment subgroup, research they contributed was verified in compliance with the PO to determine its relevance to the research paper. Material was filtered down in regard to the needs requested by the PO and ultimately included in the research paper.

SUMMARY

The deployment aspect of the research paper was covered, the challenges of bringing machine learning to edge devices. Different models used for Edge devices were discovered through research. In addition to deployment, I also looked at the monitoring aspect after deployment of the model and wrote on the future work of this project based off observations I noticed through the work I conducted.

REFERENCES

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7273223/>
- <https://cloudian.com/guides/splunk-big-data/splunk-data-analytics-splunk-enterprise-or-splunk-hunk/#:~:text=Splunk%20is%20used%20for%20monitoring,generate%20alerts%2C%20reports%20and%20visualizations.>
- <https://spacelift.io/blog/ansible-tutorial#:~:text=Ansible%20is%20one%20of%20the,is%20the%20way%20to%20go.>