

Machine Learning & Data Engineering Internship

About Orpgedia	<p>Orgpedia is a non-profit (section 8) company based in Pune whose goal is to build information extraction technologies for the Indian context (Indian languages, documents, formats). The focus of the company is to extract information from PDFs and present it in easy to consume manner so that the information trapped in the PDFs makes sense.</p> <p>More information about Orpgedia and its project can be found at http://www.orgpedia.in</p> <p>The company is formed by Mukund Deshpande as a public service initiative. (https://www.linkedin.com/in/mukunddeshpande/)</p>
Job Location	Pune – 40 % (offline) 60% (online)
Stipend Offered	None
Eligibility Criteria	<ul style="list-style-type: none">- Working knowledge of Python.- Ready to learn on-job- Innovation and ready to accept multiple technical responsibilities during the internship
Start Date/Duration	Minimum 6 months.
Project Details	<p>The project involves building a ML pipeline that takes PDF documents as input and outputs information embedded in the documents.</p> <p>The pipeline not only extracts the information, but also identifies the possible errors and in many cases corrects them.</p> <p>The pipeline uses algorithms that span across computer vision, NLP and identifies matching. In addition to the ML pipeline, there is a website that will present this data in an easy to consume manner.</p>
Required Skillset	<ul style="list-style-type: none">- Python (Basic to intermediate level)- GitHub (Basic skills)
Selection Process	Online test on Python fundamentals and 10 minute in-person discussion.
Your Contributions	<p>The entire code base is in open source (MIT licensed) and all the development will happen on GitHub.</p> <p>All your contributions would be open-source on GitHub. This will be an excellent opportunity to showcase your work.</p>