INTERNSHIP PROJECT

Remote Job Market Intelligence using Ethical Web Scraping

Presented by: Team G

Project Overview

This project focuses on collecting real remote job listings from Remote OK, building an end-to-end data science pipeline. It emphasizes **ethical web scraping** to ensure compliance with web standards, paving the way for actionable insights into skills demand and job roles.

Problem Statement & Challenges

Key Questions

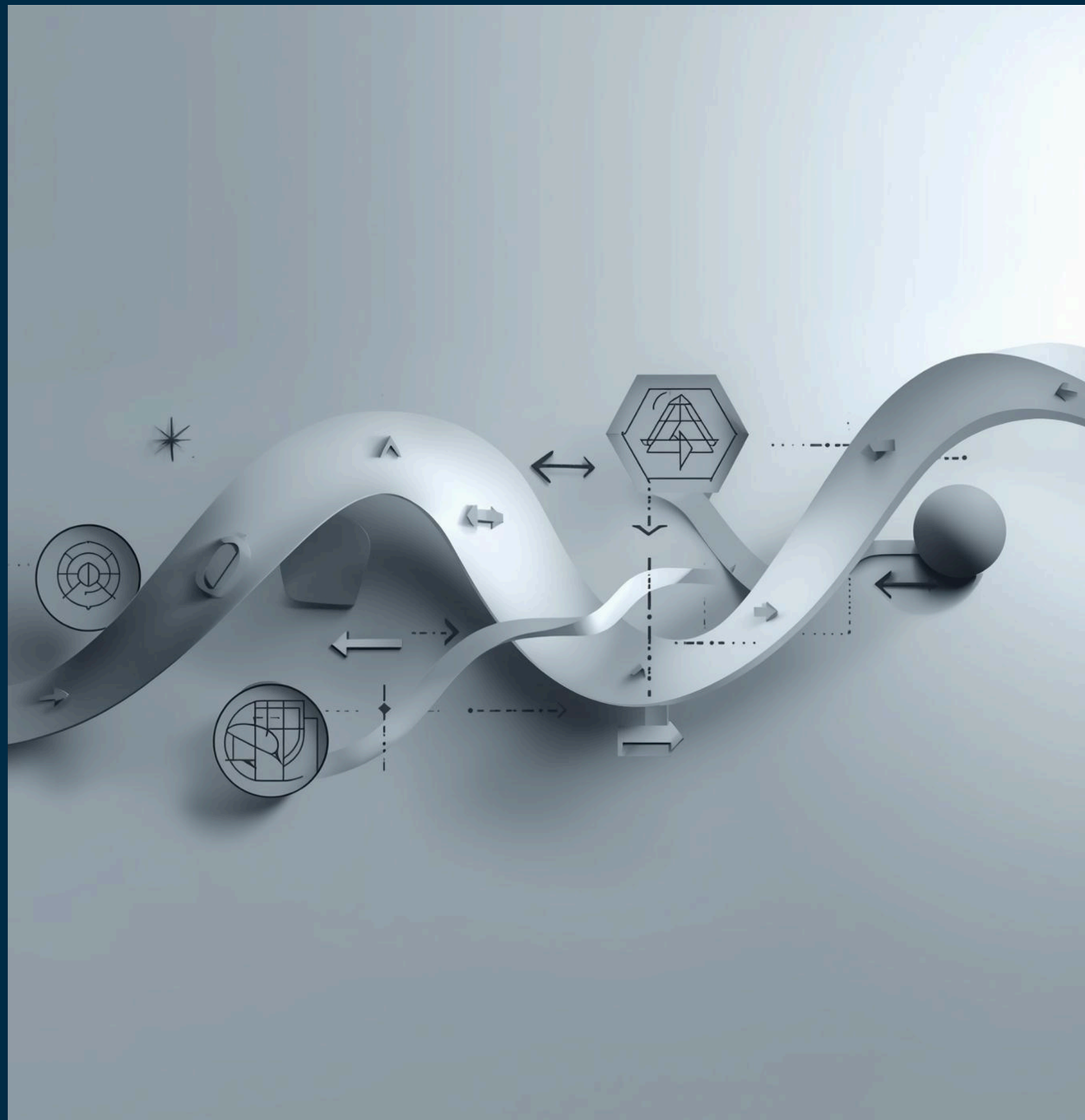
Addressing skills demand, job roles, job types, and geographic distribution in the remote job market.

Project Challenges

Managing unstructured raw data, inefficiencies of manual analysis, and adhering to ethical web scraping constraints.

Workflow Architecture

The project employs a comprehensive data science pipeline that includes stages such as website analysis, ethical web scraping, data cleaning, analysis, visualization, documentation, ethics review, and testing. Each stage is crucial and assigned to specific team members to ensure efficiency.



Website Analysis

This section outlines the **website structure mapping** conducted by **Sarang.** It details the exploration of the Remote OK site, focusing on critical HTML elements such as Job Title, Company, Skills, Location, and Job Type to ensure efficient scraping and data extraction.

Web Scraping Methodology

This section details the implementation of ethical web scraping techniques used by **Ansari and Rahul.** Utilizing Python libraries such as Requests and BeautifulSoup, the team ensured compliance with web standards while efficiently gathering job listing data from Remote OK's platform.

Data Cleaning

In this phase, we focused on refining the raw data collected from web scraping. Key tasks included **removing duplicates**, addressing missing values, and ensuring text normalization to create a clean, structured dataset that facilitates accurate analysis and insights generation. Done by **Tejas**

Data Analysis Insights

In this section, which is counducted by **Himanush** ,we present the **key findings** from our data analysis, focusing on trends in job roles, skills demand, and geographical distribution. These actionable insights provide a comprehensive understanding of the remote job market landscape.

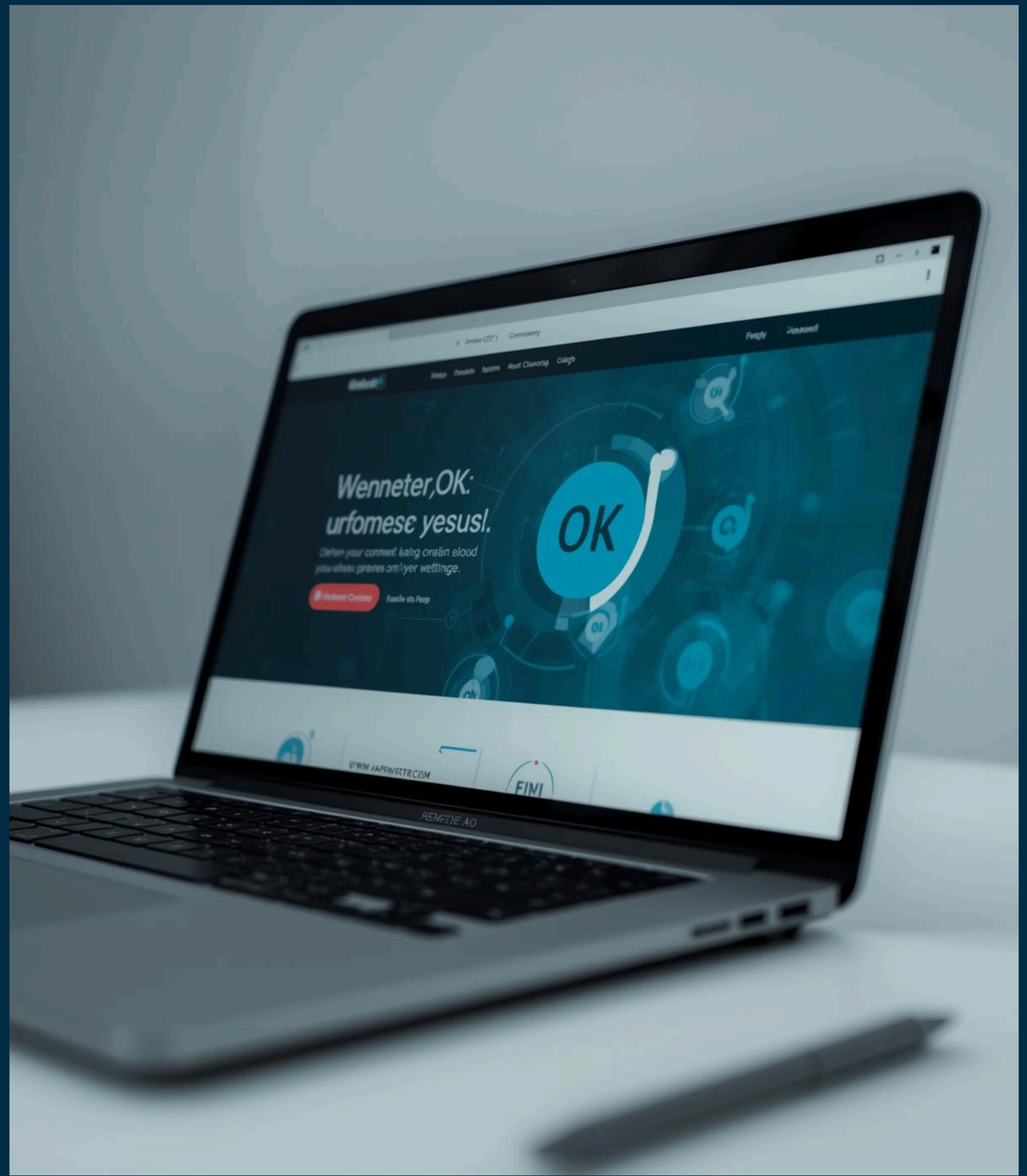
Testing & Preparation

This section outlines the final **validation process** led by **Prajakta,** ensuring the pipeline's integrity. It also highlights the collaborative efforts of **Shifa and Vimalesh** in organizing the **presentation** flow, emphasizing key insights and ethical considerations for effective communication.

Get in touch with us today!

EMAIL
hello@reallygreatsite.com

SOCIAL MEDIA
@reallygreatsite

PHONE
123-456-7890

THANK YOU