**HOW TO RUN THE RESUME CLASSIFICATION SYSTEM**

**Project:** Resume Classification System using BERT  
**Developed for:** Veridia.io AI/ML Internship  
**Submitted by:** Atharva Gaikwad  
**Date:** October 20, 2025

**Introduction**

This document provides step-by-step instructions for running the Resume Classification System on any device with internet access. The project utilizes Google Colab for execution, which provides free GPU resources and requires no local installation.

**System Requirements**

* Internet connection
* Web browser (Chrome, Firefox, Safari, or Edge)
* Google account
* Access to the provided Google Drive submission folder

**Step-by-Step Execution Instructions**

**Step 1: Access the Project Notebook**

1. Navigate to the shared Google Drive folder: Veridia\_Resume\_Classifier\_Submission
2. Open the code subfolder
3. Locate and open Resume\_Classifier.ipynb (Colab notebook)
4. The notebook will open in Google Colab automatically

**Step 2: Set Up the Streamlit Application File**

1. In Google Colab, click on the **Files** icon (folder icon) in the left sidebar
2. Create a new file by clicking the **"New file"** button or right-clicking in the file browser
3. Name the file: app.py
4. Open the app.py file from the Google Drive submission folder (code/app.py)
5. Copy all the contents from the Drive version of app.py
6. Paste the contents into the newly created app.py file in Colab
7. Press Ctrl+S (or Cmd+S on Mac) to save the file

**Step 3: Execute All Code Cells**

1. In the Colab notebook menu, click **"Runtime"** → **"Run all"**
2. Alternatively, execute each cell sequentially by pressing Shift+Enter on each cell
3. Wait for each cell to complete before the next one begins (indicated by a green checkmark)

**Step 4: Grant Google Drive Access**

1. When the code reaches the Drive mount step, a popup window will appear
2. Click **"Connect to Google Drive"**
3. Select your Google account from the list
4. Review the permissions requested
5. Click **"Allow"** to grant access
6. Wait for the confirmation message: Mounted at /content/drive

**Step 5: Upload Kaggle API Credentials**

1. When prompted with "Upload kaggle.json:", a file upload dialog will appear
2. Locate the kaggle.json file provided in the Google Drive submission folder
3. Click **"Choose Files"** and select the kaggle.json file
4. Wait for the upload to complete (a progress bar will indicate completion)
5. The dataset will automatically download and extract

**Step 6: Monitor Training Progress**

1. Observe the training progress bars and metrics displayed in the output
2. Training typically takes 14-20 minutes on Colab's GPU (Tesla T4)
3. The model will save automatically to Google Drive upon achieving the best validation score
4. Final test results will be displayed after training completes

**Step 7: Access the Streamlit Web Application**

1. After all cells complete execution, locate the public URL in the output
2. The URL will be in one of these formats:
   * ngrok: https://xxxxxxx.ngrok-free.dev
   * localtunnel: https://xxxxxxx.loca.lt
3. Copy the URL and paste it into a new browser tab
4. If using localtunnel, click **"Click to Continue"** on the warning page
5. The Streamlit application interface will load

**Step 8: Test the Application**

1. In the Streamlit interface, use the file uploader to select a resume (PDF, DOCX, or TXT format)
2. Click **"Classify Resume"** button
3. View the top-5 predicted job categories with confidence scores
4. Download results if needed using the download button

**Expected Outputs**

Upon successful execution, you will have:

* **Trained Model:** Saved in /content/drive/MyDrive/resume\_classifier\_best/
* **Visualizations:**
  + Category distribution chart (category\_distribution.png)
  + Confusion matrix (confusion\_matrix.png)
  + Training history plot (training\_history.png)
* **Artifacts:**
  + Label encoder (label\_encoder.pkl)
  + Training history (training\_history.pkl)
  + Classification report (classification\_report.txt)
* **Live Web Application:** Accessible via the provided public URL

**Performance Metrics**

* **Test Accuracy:** 88.35%
* **Weighted F1-Score:** 87.56%
* **Training Time:** ~14 minutes (GPU)
* **Categories Classified:** 24 job roles

**Troubleshooting Tips**

**Issue:** Google Drive mount fails  
**Solution:** Clear browser cookies and try again, or use an incognito window

**Issue:** Kaggle dataset download error  
**Solution:** Verify the kaggle.json file is valid and not expired

**Issue:** Out of memory error  
**Solution:** Restart runtime (Runtime → Factory reset runtime) and reduce batch size if needed

**Issue:** ngrok/localtunnel timeout  
**Solution:** The code includes automatic fallback mechanisms; wait for alternative URL to generate

**Issue:** Streamlit app not loading  
**Solution:** Ensure the Colab session is still active; the URL expires when the session ends

**Important Notes**

1. **Session Persistence:** Keep the Colab tab open while using the Streamlit app; closing it will terminate the session
2. **GPU Availability:** Colab provides limited GPU access; if unavailable, the model will train on CPU (slower)
3. **URL Validity:** The public URL changes each time you run the code; save the new URL for each session
4. **File Paths:** All file paths are pre-configured for the provided folder structure; do not modify them

**Conclusion**

By following these instructions, you can successfully run the Resume Classification System on any device. The complete end-to-end pipeline includes data preprocessing, BERT model fine-tuning, evaluation, and deployment via an interactive Streamlit web application. All artifacts are automatically saved to Google Drive for persistence across sessions.

For any technical issues or questions, please refer to the detailed README.txt and PROJECT\_REPORT.pdf included in the submission folder.