**Proposal:** Techniques, Tools, Algorithms, and Packages for a Machine Learning Model for Resume Screening

**Introduction:**

The resume screening process is an essential step in the recruitment process for many organizations. It can, however, be time-consuming and biased to screen a large number of resumes manually. This process can be automated with machine learning, resulting in a faster, more efficient, and less biased outcome. In this proposal, we outline the techniques, tools, algorithms, and packages for creating a machine learning resume screening model.

**Methodology:**

1. Collect a dataset of resumes and preprocess the data to extract features such as education, experience, skills, and certifications: using natural language processing (NPL) or names entity recognition (NER).
2. Create a job requirement list for different hiring job positions in the company (10 positions)
3. Split the data into training (80%) and validation sets (20%):
   * Qualified resumes: resumes from current or past employees in such job roles
   * Non-qualified resumes: all the other resumes from other job roles
4. Choose a suitable algorithm for classification such as logistic regression, support vector machines (SVM), or neural networks.
5. Train the model: Train the model on the training dataset using the selected algorithm. Transfer learning could be applied in this final project where a pre-trained model is used as the starting point for a new model. A pre-trained NLP model, such as BERT or GPT-3, can be used as the starting point for the new model.
6. Evaluate the model: Evaluate the model's performance on the validation dataset using metrics such as accuracy, precision, recall, and F1 score.
7. Deploy the model: Once the model's performance is satisfactory, deploy it in production.

**Tools:**

Python: Python is a popular programming language for machine learning and has many libraries for NLP, such as NLTK, spaCy, and TextBlob.

Jupyter Notebook: Jupyter Notebook is an open-source web application that allows for interactive coding and data visualization.

GitHub: GitHub is a web-based platform for version control and collaboration. It can be used to store the code and models for the machine learning project.

**Algorithms:**

Logistic Regression: Logistic regression is a linear model used for classification problems, such as the binary classification problem of resume screening.

Support Vector Machines (SVMs): SVMs are another popular algorithm for classification problems. They work by finding the hyperplane that maximally separates the two classes.

Random Forest: Random Forest is an ensemble algorithm that combines multiple decision trees to improve the accuracy of the model.

**Packages:**

Scikit-learn: Scikit-learn is a Python library for machine learning that includes implementations of many algorithms, including logistic regression, SVMs, and random forest.

 Transformers: Transformers is a Python library for state-of-the-art NLP models, such as BERT and GPT-3.

 Pandas: Pandas is a Python library for data manipulation and analysis. It can be used to preprocess and clean the resume data before feeding it into the machine learning model.

**Conclusion:**

Data will be gathered by from LinkedIn save under .csv format. Using NLP, supervised learning, and transfer learning, we can construct a machine learning model for resume screening. Jupyter Notebook, Python, and GitHub are useful tools for developing and storing the code. Various algorithms can be used to classify resumes, such as logistic regression, SVMs, and random forests. Data can be manipulated and analyzed with a variety of libraries, including Scikit-learn, Transformers, and Pandas.