明白了，你打算使用你自己的情感分类模型来预测歌手写歌的情感和听众喜好的情感。在这种情况下，你可以按照以下步骤进行分析：

1. **数据获取：** 获取包括歌手写歌的情感标签和排行榜歌曲的情感标签的数据，确保数据中包含年代信息。
2. **数据整合：** 将歌手写歌的情感数据和排行榜歌曲的情感数据整合到一个数据集中，按照年代进行组织。
3. **年代分组：** 根据不同的年代对数据进行分组，以便分析不同时期的情感变化。
4. **模型预测：** 使用你的情感分类模型对歌手写歌的歌曲和排行榜歌曲进行情感预测。模型将为每首歌曲分配情感标签。
5. **分析歌手情感：** 分析不同年代歌手写歌的情感趋势，基于模型的预测结果。你可以查看每个年代歌手写歌的主要情感趋势。
6. **分析听众情感：** 分析不同年代排行榜歌曲的情感，基于模型的预测结果，以了解不同时期听众喜好的音乐情感。
7. **对比和关联：** 尝试对比不同年代歌手写歌的情感趋势和听众情感趋势。确定是否有关联或影响。
8. **可视化和报告：** 使用图表和可视化工具来呈现你的分析结果。撰写一份报告或文章，详细描述你的研究和发现。

通过使用你的自己的情感分类模型，你可以根据自己的方法对音乐情感进行预测和分析。这将使你有机会验证和比较你的模型在预测不同年代音乐情感方面的性能，并深入探讨歌手和听众之间的情感互动。

CBF作为基线来进行比较

* **时代性情感趋势**：
  + 通过年份或特定时间段，观察排行榜上歌曲情感的分布变化。例如，某些时期可能偏向高唤醒积极的歌曲，而其他时期可能偏向低唤醒消极的歌曲。
* **流派与情感的关系**：
  + 分析不同的音乐流派在四个情感区域的分布。例如，电子舞曲可能更偏向高唤醒积极，而乡村音乐可能更偏向低唤醒积极。
* **特定事件与情感反应**：
  + 分析特定事件后，如社会或政治事件，排行榜上的歌曲情感是否有所改变。
* **艺人的情感特色**：
  + 对于常上排行榜的艺人，分析其歌曲在四个情感区域的分布，看看是否有某个或某几个艺人有特定的情感特色。
* **情感持续性与排行榜位置**：
  + 分析在排行榜上位置靠前或停留时间长的歌曲的情感标签，看看是否某一情感类别的歌曲更受欢迎。
* **地区与文化差异**：
  + 用不同地区的排行榜数据，分析不同地区的情感偏好是否有所不同。
* **情感与其他歌曲特性的关系**：
  + 结合歌曲的其他属性（如节奏、旋律、时长等），看看是否某些情感标签与特定的歌曲属性有关联。
* **听众反馈与情感关系**：
  + 如果排行榜提供了听众的评价或反馈，分析这些反馈与歌曲的情感标签是否有关联。

1. **CBF (Content-Based Features)**：基于内容的特征，主要是词袋模型的特征。这包括不同的n-gram（如1-gram到3-gram）和POS tagger应用后的n-gram。这些特征作为大多数研究的基准。
2. **StyBF (Stylistic-Based Features)**：基于风格的特征。这些特征考虑了歌词的文体特点。例如，使用POS tagger统计的歌词中不同的语法类别或标签的数量，以及新颖特征如#Slang（俚语词数量）和与大写字母相关的特征（如ACL和FCL）。
3. **StruBF (Structural-Based Features)**：基于结构的特征。这些特征考察歌词的结构组成，例如歌词中的诗句结构、节奏等。这个特征集只包含新颖特征。
4. **SemBF (Semantic-Based Features)**：基于语义的特征。这些特征主要考虑歌词的意义和情感倾向。包括从不同情感词典和工具（如Synesketch、ConceptNet、LIWC、GI）中提取的特征，以及新颖特征如gazetteers、DAL和ANEW中的特征。

通过这些不同的维度，他们能够更全面地捕获歌词中的情感信息，从而提高情感分类的准确性。

1. **排行榜的歌曲数据集（代表听众的喜好）**:
   * **来源**：您可以从流行的音乐排行榜（如Billboard、iTunes、Spotify等）中收集数据。
   * **选择方法**：选择特定的时间段（如每年的年终排行榜）的前N首歌曲。这些歌曲基本上代表了在那个时间段最受听众喜欢的歌曲。
   * **目的**：分析这些歌曲的歌词，以了解在特定时间段内，听众更喜欢哪种情感或歌词风格。
2. **歌手的歌词数据集**:

* **来源**：这个稍微复杂一些，因为“歌手的歌词”是一个宽泛的概念。这里有几种可能的选择：
* a. **特定的歌手或团体**：选择某些具有代表性的歌手或团体，收集他们的全部或部分歌词。
* b. **新晋歌手**：您可以选择每年新出道的歌手的歌曲，以了解新晋歌手的歌词风格和情感。
  + c. **非排行榜歌手**：选择那些没有进入主流排行榜但在某些领域有影响力的歌手或团体。
  + **目的**：了解歌手或创作者如何表达自己的情感，以及他们的歌词风格和情感如何随时间变化。

对于这两个数据集，您需要做的是：

1. **数据收集**：使用网页抓取工具或API从音乐网站、数据库或其他来源收集歌词。
2. **数据预处理**：清洗数据，如去除重复歌词、纠正拼写错误等，并为每首歌曲添加标签（如歌手、年份、是否上榜等）。
3. **模型应用**：将您的情感识别模型应用于这两个数据集，提取每首歌曲的情感或风格特征。
4. **分析比较**：根据提取的特征分析和比较这两个数据集，找出任何可能的趋势或模式。

通过以上步骤，您可以得到关于听众喜好和歌手歌词风格的洞察。