《Classifying Question Papers with Bloom's Taxonomy Using Machine Learning Techniques》（2019，International Conference on Advances in Computing and Data Sciences计算和数据科学进展国际会议）

论文中对题目进行布鲁姆分类层次划分的方法，核心上是一套**有监督的机器学习流程**，可分为以下几个步骤：

1. **数据标注**
   作者首先由教育专家按照布鲁姆分类法的六个认知层级（记忆、理解、应用、分析、评价、创造）对 1,024 道试题进行人工标注，生成用于训练与测试的带标签数据集 。
2. **文本预处理**
   - **分词**：利用 NLTK 的单词分词器将每道题切分成词项列表。
   - **去停用词**：准备停用词表（如 “the”、“is” 等无意义高频词），并将其从分词结果中过滤掉 。
3. **特征提取**
   - **布鲁姆动词特征**：根据各层级的典型动作动词（如 “define”、“analyze”、“create” 等），提前构建“先验”动词特征列表。
   - **词频向量化**：对所有题目文本应用 Count Vectorizer，构建词袋模型的词频向量。
   - **特征组合与筛选**：将动词特征与词频特征合并后，人工从中挑选出最能区分各层级的显著特征，形成最终的输入特征集。
4. **模型训练与调优**
   - 将数据集划分为训练集与测试集后，分别在七种分类器（KNN、决策树、随机森林、SVM、神经网络、LDA、逻辑回归）上进行训练。
   - 通过网格搜索（Grid Search）系统地调节各模型超参数（如 KNN 的邻居数、SVM 的 CCC 值、神经网络的 epochs / batch size 等），并结合交叉验证（Cross-Validation）选取最优配置。
5. **多分类输出**
   最终，每道试题被分到六个层级之一。模型性能评估表明，线性判别分析（LDA）与逻辑回归（LR）在这一多类分类任务上以 83.3% 的准确率领先

EXAM QUESTIONS CLASSIFICATION BASED ON BLOOM'S TAXONOMYCOGNITIVE LEVELUSING CLASSIFIERS COMBINATION （2015，Journal of Theoretical and Applied Information Technology理论与应用信息技术杂志）

提出了一种**有监督机器学习 + 投票融合**的方法来自动对试题进行分层，其核心流程如下：

1. **数据准备与人工标注**
   - 从英国国民大学（UKM）2006–2011 年计算机编程考试题库中收集短文本试题，并由教学领域专家依据布鲁姆分类法的六个认知层级对每道试题进行手工标注，构建带标签的数据集 。

2. **文本预处理**
   - **分词（Tokenization）**：按空格和标点将每道试题划分为词项。
   - **去除停用词**：删除非字母字符和诸如“the”、“is”之类的无意义高频词，以减少噪声 。
3. **特征选择（Feature Selection）**
   - 使用三种统计量从词袋向量中筛选最具判别力的特征：
     - **互信息（Mutual Information, MI）**：衡量词项与层级标签间的依赖关系；
     - **卡方检验（Chi-Square）**：评估词项出现与否与分类标签间的关联度；
     - **比值比（Odd Ratio）**：比较词项在某一层级与非该层级中的分布差异 。
4. **基分类器训练**
   - 在保留和不保留特征选择的情况下，分别用以下三种经典分类器对试题进行训练：
     1. **支持向量机（SVM）**
     2. **朴素贝叶斯（NB）**
     3. **k-近邻（k-NN）**
5. **投票融合（Voting Algorithm）**
   - 对每道试题，收集三种分类器的预测结果，以“多数票”原则决定最终的布鲁姆层级归属。该方法能够综合各模型优势，纠正单一分类器可能出现的误判。

通过上述“**标注 → 清洗 → 特征筛选 → 多模型训练 → 投票决策**”流水线，论文实现了对短文本考试题目按布鲁姆认知层级的自动分层，最终在使用互信息筛选 250 个特征并融合三模型时，达到了最高的宏 F_1 评分约 92.28% 。

Automatic Classification of Learning Objectives Based on Bloom's Taxonomy(2022年，Proceedings of the 15th International Conference on Educational Data Mining

第 15 届教育数据挖掘国际会议论文集)
作者提出了一套端到端的、基于监督学习与深度学习相结合的方法来将学习目标（文本“题目”）映射到布鲁姆认知领域的六个层级（记忆、理解、应用、分析、评价、创造）。其核心流程可分为以下几步：

1. **数据收集与人工标注**
   - 从澳大利亚一所大学的 5,558 门课程中抓取了 21,380 条学习目标，并由受过布鲁姆分类法培训的人工编码者进行标注，每条学习目标可对应一个或多个认知层级 。
2. **文本预处理**
   - **全部小写**：将文本统一转为小写以降低词表稀疏性。
   - **停用词过滤**：在构造 n-gram 和 TF-IDF 特征时剔除英语高频无意义词。
3. **特征工程**
   - **传统机器学习特征**（共 3,094 维）
     - 1,000 最常见的**单词**与 1,000 最常见的**双词短语**（均排除停用词）
     - 基于上述词表计算的**TF-IDF** 特征（1,000 维）
     - **自动可读性指数**（Automated Readability Index）
     - **LIWC 词典**衍生的 93 个心理语言特征（如认知过程词、功能词等）

- **深度学习特征**
  - 直接使用预训练的 BERT-uncased 模型来生成上下文敏感的句子级嵌入，不再额外提取手工特征。

4. **分类器构建**
   - **二元分类器方案**：为布鲁姆的每个认知层级训练一个二分类器（label vs. non-label），共计 6 × 6 = 36 个二元模型。
   - **多类多标签方案**：训练单一的多类多标签模型，让它同时预测所有可能的层级标签。
   - 在五种传统算法（朴素贝叶斯、逻辑回归、支持向量机、随机森林、XGBoost）和基于 BERT 的深度模型上，作者均尝试了上述两种方案 Li 等 - 2022 - Automatic…Li 等 - 2022 - Automatic…。
5. **超参数调优与训练**
   - **机器学习模型**：在 80% 的训练集上使用 3-折交叉验证和网格搜索来选择最优参数，评价指标为 F1 分数。
   - **BERT 模型**：对下游分类层和所有 BERT 层同时进行微调，批量大小设为 64，训练 3 个 epoch，并启用早停，当验证集 F1 连续 10 次不增时终止训练。
6. **性能评估**
   - 在留出的 20% 测试集上，分别计算准确率、Cohen’s κ、AUC 与 F1 分数，比较二元与多类多标签模型，以及不同算法之间的表现。

通过上述流程，Li 等人证明了：

- **BERT-based 二元分类器**在所有层级上表现最佳（κ 达到 0.93、F1 达到 0.95）。
- **SVM、随机森林、XGBoost** 等传统模型在大规模数据下也能达到接近 BERT 的高性能。
- 相比于一次性预测所有层级，多条二元分类器方案略微优于多类多标签模型

An effective deep learning pipeline for improved question classification into bloom’s taxonomy’s domains（2022，SCI教育学二区）

在 Sharma 等人（2023）所提的流水线中，对试题进行布鲁姆认知层级分类的核心方法可归纳如下：

1. **数据准备与预处理**
   - 收集并合并多源数据集（Yahya 等, Mohammed & Omar, TREC 及自建数据集），对每道题目人工标注六个认知层级标签（记忆、理解、应用、分析、综合、评价）。
   - 文本清洗包括统一小写、去除标点与无意义符号、剔除过短（<2 字符）或过长（>21 字符）单词，并按 80 : 20 划分训练 / 测试集 。
2. **序列化与向量化**
   - 使用 TensorFlow/Keras Tokenizer 将清洗后的题目转换为整数索引序列，设定词汇表大小 1 692、最大序列长度 40，超出部分截断，较短序列后填充 0；
   - 目标标签采用 One-Hot 编码为长度 6 的向量 。
3. **深度学习分类器架构**
   - **Sequential 模型**：

    1. **Embedding 层**（随机初始化或预加载 GloVe/ELMo 向量）→
    2. **特征抽取层**（Conv1D + （双向）LSTM）→
    3. **Dense 输出层**（6 个神经元 + Softmax）
  - **功能式模型（Functional API）**：
    - **BERT-Based**：输入 Token IDs、Mask IDs、Segment IDs，经 KerasLayer 封装的预训练 BERT（可微调）→ 池化输出 → 若干全连接层 → Softmax 分类。
  - **ELMo-Based**：通过 TensorFlow Hub 加载预训练 ELMo 模块生成上下文词向量，再经 Conv1D + BiLSTM → Dense → Softmax。
4. **训练与优化**
  - 优化器均采用 Adam（BERT 模型学习率约 1e-5，其它模型可设 1e-3–1e-4），损失函数为多类交叉熵，常用 Dropout、梯度裁剪等正则化手段防止过拟合；
  - 训练轮数 5–10 轮，批量大小 8–16，通过验证集监控早停 。

最终，在留出的测试集上，BERT + Dense 网络实现了约 81.1% 的分类准确率，优于其他基于 Conv1D + BiLSTM 及 ELMo 的方案。

Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec（）

针对多领域的问答短文本，提出了如下基于布鲁姆认知领域对试题分层（分类）的方法：

1. **问题收集与标注**
   从两个开放领域数据集中共收集 141 + 600 道开放式试题，并由专家依照布鲁姆认知领域的六个层级（记忆/记忆、理解、应用、分析、综合、评价）为每道试题打标签，构建带标签的数据集 。
2. **文本预处理**
  - **归一化**：去除标点、数字与非英文字符；统一小写；保留部分能区分层级的停用词（如 what、how、your 等）。
  - **分词与词性标注**：按空格切分为 token，使用 Stanford POS Tagger 打上词性标签（VB, NN, JJ, RB 等）。
  - **词干提取**：对计算 TFPOS-IDF 特征时使用 Porter 词干器，但不对 word2vec 特征做词干，以保证嵌入向量一致性 。
3. **特征提取**
  - **TFPOS-IDF（基于词性加权的 TF-IDF）**
    按照实验验证，给**动词**赋予最高权重 w1w_1w1、名词/形容词次高 w2w_2w2、其余词最低 w3w_3w3；在计算局部词频时，将每个词的出现次数乘以其词性权重，再归一化后乘以 IDF，得到加权稀疏向量表示 。
  - **预训练 Word2vec 嵌入**
    利用 Google News 预训练的 300 维 word2vec 模型，将每个词转换为 300 维实数向量；若词不在词表中，则用全零向量代替 。

- **特征融合（W2V-TFPOS-IDF）**
  对于文档（试题）中每个词，对所有词向量求和，得到该试题的单一300维密集向量。

4. **分类器训练与多层级分类**
   将上述向量输入到三种常用监督分类器——KNN、逻辑回归（LR）、支持向量机（SVM），分别训练并评估，每道试题最终被分入布鲁姆的六个认知层级之一。