**“东证期货杯”全国大学生统计建模大赛选题**

**选题1: 基于互联网公开数据的企业认定、评价、筛选模型的构建**

对企业进行科学的行业分类是研究企业风险、评价企业发展趋势的基础，在互联网信息爆炸增长的今天，如何有效运用互联网公开信息对企业进行科学的评价、认定、筛选成为学界、业界各方关注的焦点。

传统的企业分析主要基于财务数据进行开展，在产业持续转型和创业创新企业不断涌现的今天，传统的分析方法难以客观准确的对目标企业做出合理的评价。相较于传统的分析方法，基于互联网大数据与人工智能的分析方法能够从多维度海量企业信息中挖掘出有助于科学、量化分析企业的信息，因此基于大数据和人工智能的企业分析方法的构建具有较强的实践意义。

企业的真实发展情况往往难以准确及时了解，特别是对于创新型科技创业企业而言，企业往往轻资产化，发展年限短，企业真实信息更加难以获取。但对于政府、金融机构、创投机构，及时、准确了解企业具有重要意义。根据企业公开信息，构建企业的评价体系，多维度分析、量化评价企业（例：企业的运营情况、发展能力、创新能力、投资前景、负面新闻、舆情指数等）建立企业画像。并从创投机构的角度出发，量化创投机构投资偏好，为创投机构推荐适合的优秀创业企业，为创业企业寻找适合的创投机构融资。

对于本题，报名期结束后，组委会将统一发放有关备选指标数据及行业分类标准的相关信息，参赛者需要根据提示自助获取企业样本原始数据，运用文本挖掘等方法，完成对样本企业行业分类模型的构建及样本外检测，并最终提交PDF版本的初赛论文正文、数据分析程序及原始数据。

**选题2: 互联网金融时代下信用评分体系模型的构建**

当下互联网金融已蓬勃兴起，呈现出多种多样的业务模式和运行机制。金融机构能够突破时间和地域的约束，在互联网上为有融资需求的客户提供更快捷的金融服务。通过互联网技术，加快业务处理速度，带给用户更好的服务体验。但同时存在着信用风险和用户欺诈等问题，急需通过信用评分模型提高风险控制水平。

征信机构利用采集到的丰富信息对个人进行综合信用评价。在丰富海量的个人信用历史和信用行为数据基础上，采用数据挖掘方法得出的信用行为模式能够更加准确地预测个人未来的信用表现，能够提高操作的效率，降低授信成本，精确估计消费信贷的风险，是金融机构内部评分不可替代的重要工具。因此，建立精准的信用评分体系对于企业有着重要的意义。

对于本题，报名期结束后，组委会将统一发放某贷款机构的历史业务数据作为原始数据，参赛者需要基于给定数据，运用数据挖掘等方法，构造模型变量，制定信用规则，建立信用评估模型，预测违约情况，并最终提交PDF版本的初赛论文正文及数据分析程序。

**选题3:** **基于高频数据的股票量化交易策略的构建**

高频数据和超高频数据分析是现代金融领域最前沿的热点问题之一。随着信息技术的不断革新，电子交易平台在国际证券市场上被逐步推广使用，这使得金融市场的资源配置效率得到了提高。更重要的是，与此同时人们能方便地从市场上获取和存储高频交易数据，这些海量交易信息将比以往的低频数据更能帮助我们去认识市场本质，改善市场机制设计，进而有效应对金融风险、实现市场的资源配置效率、促进分工发展的职能。然而，像买卖差价、价格离散取值等市场微观结构效应在高频数据下被凸显出来，这些都给高频数据的计量建模和分析带来了诸多前所未有的挑战。

准确的预测和分析最终是为交易服务的，本题将基于股票高频数据，设计出有效可行的量化交易策略。量化交易作为证券期货投资交易管理业的核心与基本交易工具，目前已被国际投资管理业基金经理们普遍采用，是现代投资研究领域的重点发展方向。综观中国证券和衍生品市场，量化交易策略的研究和应用均处在起步阶段。我们希望通过对本题的研究，促进参赛者对于股票交易自动化、智能化、投资决策科学化的理解，最终达到风险控制和投资收益最大化的目标。因此，量化交易策略的构建具有重要的科学价值与实践意义。

对于本题，报名期结束后，组委会将统一发放某证券机构的历史交易五分钟高频数据，参赛者需要基于给定数据，运用金融统计、机器学习等方法，完成股票量化交易策略的构建，从风险和收益两个方面进行分析，并最终提交PDF版本的初赛论文正文及数据分析程序。