

业务数据；外部信息包括各类教育信息、外部统计和调研数据及文档等。

(二) 数据存储与管理

数据存储与管理层是整个系统的核心，包括 ETL 管理工具、公共数据集、元数据、数据仓库和数据集市。高校 BI 系统平台建设采用数据驱动设计方法，从学校原有的各个部门的业务处理系统和外部数据源中经过 ETL 提取数据，并根据常见的分析和统计主题，建设校级数据仓库以及人才培养、师资队伍、科学研究、办学资源、交流合作等主题的数据集市。

(三) 数据查询与分析

高校决策者常常希望从不同的角度审视教育数据，比如从时间、区域、学科、教学或科研成果、课程建设、学生层次、交流合作、办学资源等维度全面了解学校的教育质量和状态。高校 BI 系统平台的数据分析层利用商业智能技术为高校管理主要提供固定报表、即席查询、统计分析、多维分析、预警功能、预测分析、数据挖掘建模分析及优化分析等，根据学校现有学生、教师、资源、科研和人才培养等状况，有助于高校决策者全面地对学校资源配置进行调控、对学校整体办学信息的内部结构进行调整等，做出对学校发展更为有利的科学决策。

(四) 用户接口

用户接口层根据高校用户访问需求和角色访问授权机制，提供强大的多用户数据查询操作，并以仪表盘或表格、直方图、饼图等直观方式将查询结果或决策信息呈现给用户。

三、应用案例

下面以高校 BI 系统平台中的调研数据为商业智能技术应用案例，利用回归方法对大学生学习成果进行数据挖掘分析。

(一) 数据来源

案例分析的数据来源于高校 BI 系统平台中“中山大学学生学习状况调查”项目于 2012 在中山大学全校范围内开展的在线调研数据。^[13]调查覆盖全校 36 个学院（系），调查总体约为 3.3 万名本科生。让学生在无压力的情况下答题，共回收问卷 7051 份，回收率约为 21.2%，与国际上通用的问卷回收率相当。案例分析聚焦于本科样本，全部回收的问卷根据答题时长、问卷质量标准等原则，筛选出有效问卷数据 6673 份，有效率为 94.6%。

本研究从学生学习经历角度，在“生源－学习－成果”的逻辑框架中，考察分析学校因素和学生因素对于学生学习成果的影响机制。调查把学生学习经历和成果分解为生源情况、学校学习资源供给、学生与学校的融合、学生学习投入、学生成果、学校成果 6 大维度，各维度下题目的内部一致性均达到 0.9 以上，具有较高的信度。

(二) 数据分析

逐步回归提供了一种识别与学生学习成果相关的具体经历的方法，对于学生学习状况调查中的 227 项进行相似项合并，用向前和向后逐步回归确定与学习成果相关的项目，对残差图和诊断法的彻底审查，最后确定 17 个独立变量出现在多元回归模型中（如表 2 所示），其中，相关系数 R 为 0.994，校正判定系数 R^2 为 0.988，因变量变化中有 98.8% 左右的信息可以由预测变量解释，说明模型的拟合优度较好；Durbin－Watson 为 1.937，接近最佳理想值，如表 1 所示。

表 1 学习成果的回归分析模型

模型	R	R 方	调整 R 方	标准估计的误差	Durbin-Watson
1	.994	.988	.988	12.19341	1.937

表 2 显示的是回归系数的相关统计量，可以看出，这 17 个独立变量的显著性概率 Sig. 都小于 0.05，说明其系数显著不为 0，这 17 个变量均与学生学习成果显著相关。

表 2 回归系数的相关统计量

	非标准化系数		标准系数	t	Sig.
	B	标准误差			
学校学习资源提供					
课程作业评价	.503	.052	.068	9.664	.000
专业学习经历评价	.184	.037	.036	4.953	.000
学术规范指导	.084	.038	.010	2.218	.027
学生学习投入					
自主学习情况	.577	.030	.165	18.953	.000
活动参与情况	.165	.042	.028	3.967	.000
课外阅读情况	.475	.073	.027	6.529	.000
论文写作情况	.301	.104	.007	2.903	.004
朋辈交流情况	.603	.030	.185	19.877	.000
讨论关注的内容情况	.838	.067	.111	12.519	.000
师生交流	.250	.033	.038	7.624	.000
课业活动	.375	.042	.036	8.956	.000
个人闲暇活动时间分配	.177	.025	.042	6.993	.000