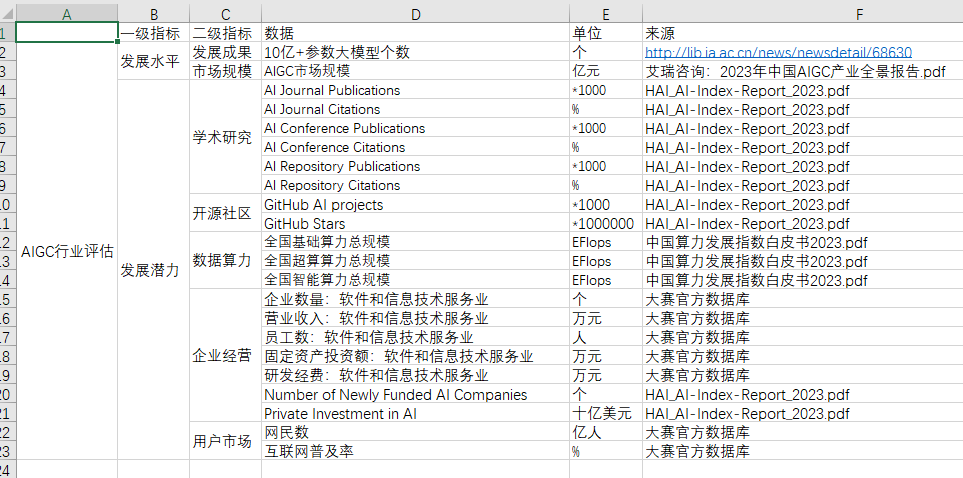
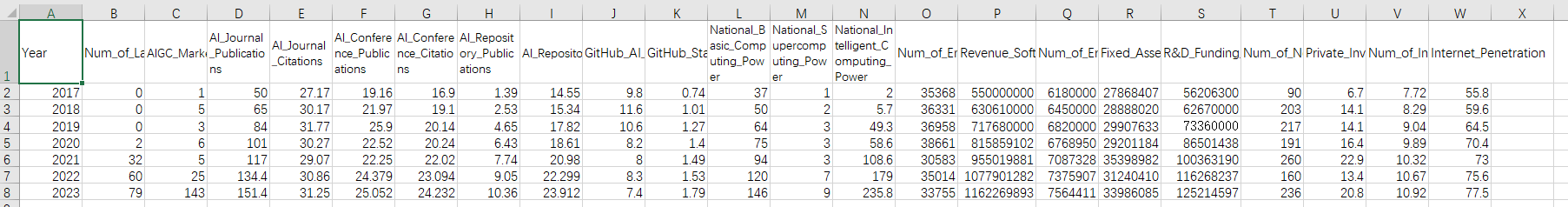
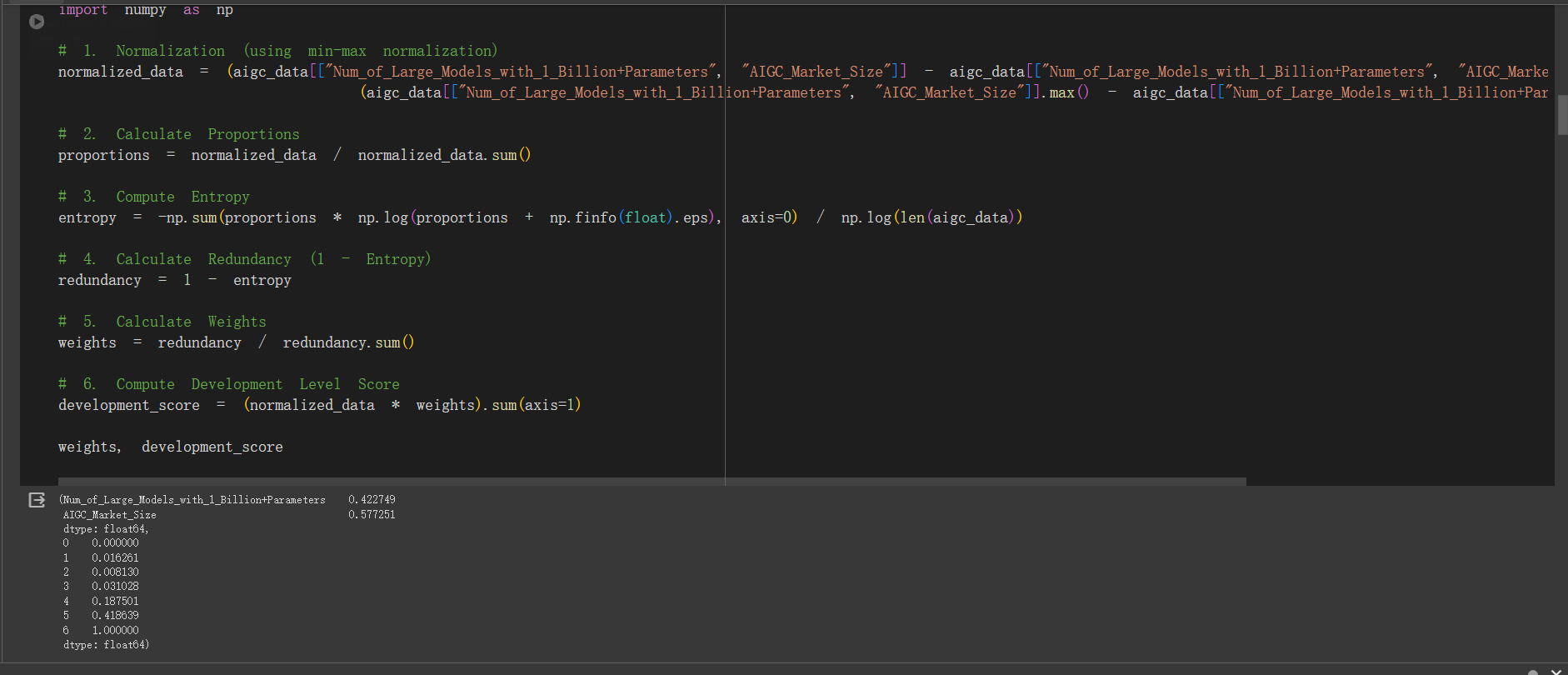
**设计指标**

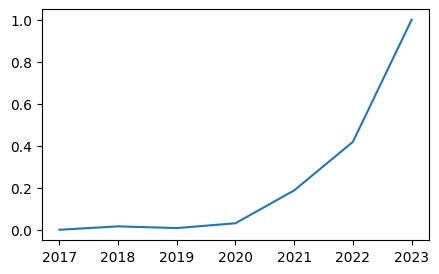
****

**收集数据（基本只能找到年度数据，而且对新兴行业而言，太老的数据似乎价值也不大？）**

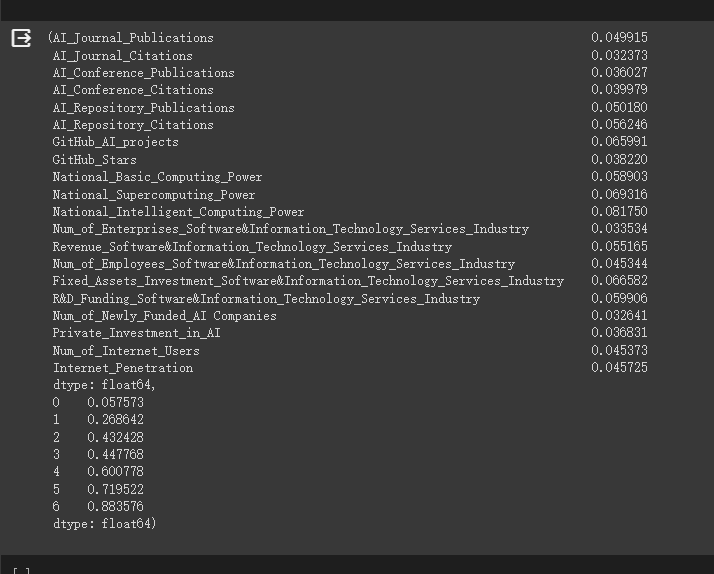
****

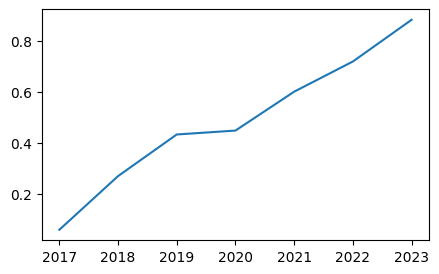
**对所有数据归一化。熵权法求模型个数和市场规模的权值，加权求和，得到发展水平得分**

****

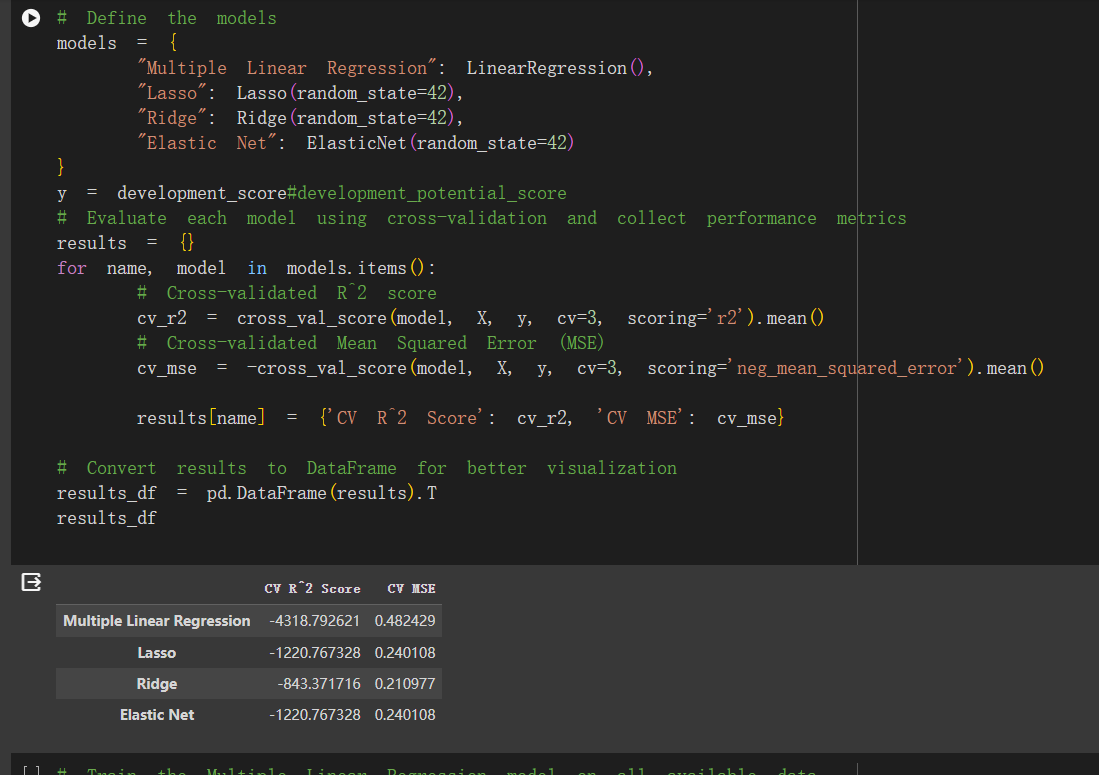
****

**熵权法，对其他指标加权求和，得到发展潜力得分**

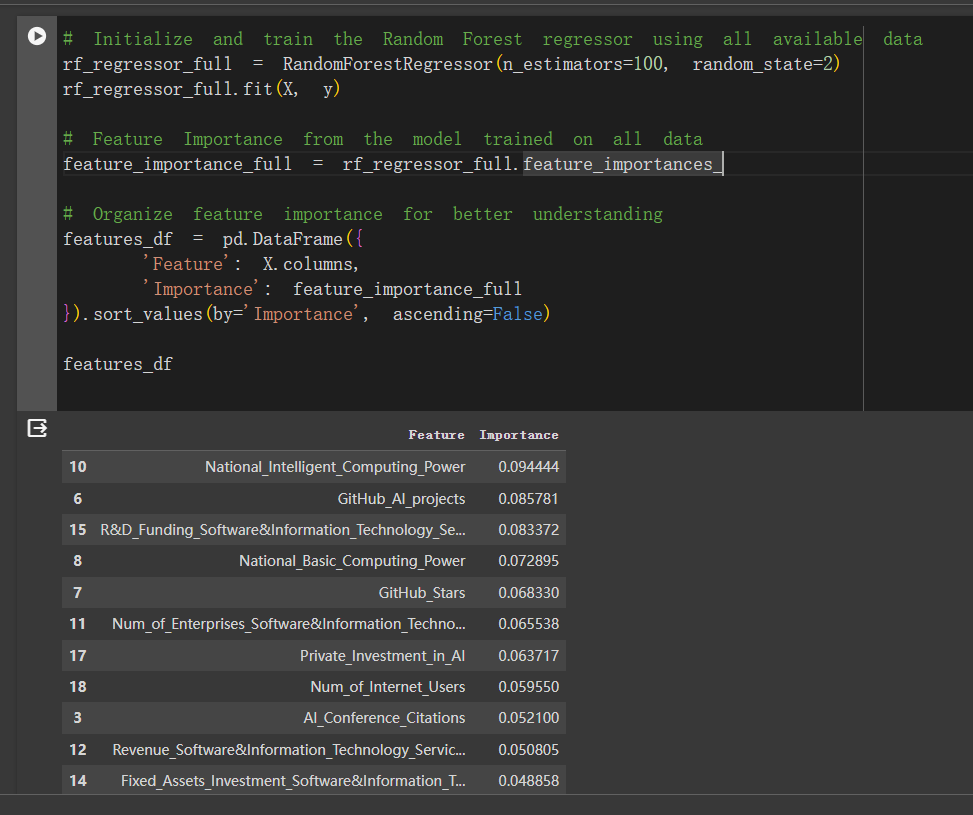
****

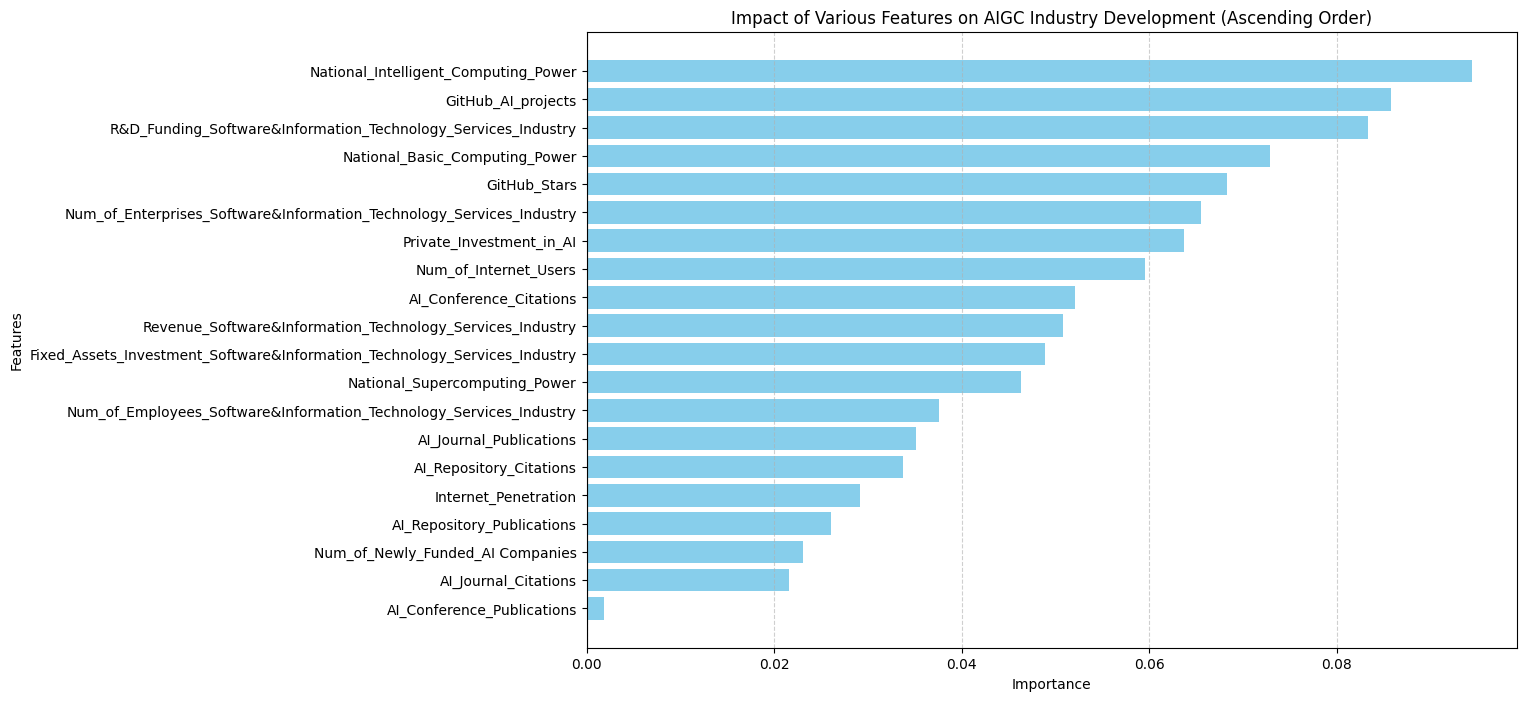
****

**用线性、拉索、岭、弹性网回归模型拟合各个指标与发展水平得分的关系，似乎效果都不佳**

****

**用随机森林****拟合各个指标与发展水平得分的关系，数据全部作为训练集，****得到各个指标的重要程度，也在一定程度上验证了熵权法给出的权重结果**

****

****

**初步结论：重点发展算力，尤其是智能算力；促进开源社区发展（但GitHub不能完全代表中国开源社区水平）；鼓励软件行业研发投资……**