Class Distribution (by file count)

- business: 22.9%
- tech: 18.0%
- sport: 23.0%
- politics: 18.7%
- entertainment: 17.3%