BAB V
PENUTUP

5.1. Kesimpulan

Berdasarkan hasil pengujian terhadap *benchmarking framework* DVWA yang dijalankan secara daring (VPS), perbandingan antar *Web Vulnerability Scanner* (WVS) menunjukkan bahwa tidak terdapat satu pun alat yang unggul secara menyeluruh pada seluruh metrik evaluasi. Secara umum, *precision* pada sebagian besar alat menunjukkan nilai tinggi, yang mengindikasikan rendahnya *false positive* dalam lingkungan terkontrol. Namun, nilai *recall* dan *F1-score* bervariasi signifikan antar alat. Xray menunjukkan *recall* yang lebih tinggi pada kategori kerentanan berbasis eksploitasi, sedangkan ZAP dan Nikto menunjukkan *recall* yang lebih tinggi pada kategori kesalahan konfigurasi (*misconfiguration*). Dari sisi waktu eksekusi, terdapat variasi yang dipengaruhi oleh mekanisme *crawling*, strategi pemindaian, serta sensitivitas terhadap latensi jaringan, di mana Nuclei menunjukkan penurunan performa pada VPS akibat *timeout*, tetapi meningkat ketika dijalankan secara lokal.

Jika ditinjau berdasarkan OWASP Top Ten 2025, masing-masing alat memperlihatkan pola spesialisasi yang jelas. Xray lebih efektif dalam mendeteksi kerentanan pada kategori A05 (*Injection*) dan A07 (*Authentication Failures*), yang memerlukan pendekatan *active scanning* dan verifikasi eksploitasi. Sebaliknya, ZAP dan Nikto lebih konsisten dalam mendeteksi kerentanan pada kategori A02 (*Security Misconfiguration*), seperti *missing security headers* dan *directory browsing*. Pola ini menunjukkan bahwa pendekatan teknis alat, baik berbasis eksploitasi aktif maupun berbasis analisis konfigurasi sangat memengaruhi tingkat *recall* terhadap jenis kerentanan tertentu.

5.2. Saran

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya maupun implementasi praktis di lingkungan organisasi.

Pertama, penelitian selanjutnya disarankan untuk melakukan pengujian tambahan menggunakan *Web Vulnerability Scanner* (WVS) komersial secara langsung pada lingkungan dan *ground truth* yang identik. Hal ini bertujuan untuk memperoleh perbandingan numerik yang lebih adil dan terkontrol, khususnya dalam mengukur metrik *recall* dan *precision* pada kategori kerentanan yang sama. Dengan demikian, perbandingan

antara alat *open-source* dan komersial dapat dilakukan secara empiris tanpa bergantung sepenuhnya pada studi terdahulu yang mungkin memiliki perbedaan metodologi.

Kedua, penelitian mendatang dapat memperluas cakupan *framework* uji dengan menggunakan benchmark tambahan seperti OWASP Benchmark atau aplikasi web berbasis arsitektur modern (misalnya REST API atau *Single Page Application*). Hal ini penting karena karakteristik kerentanan pada aplikasi modern dapat berbeda dari aplikasi berbasis PHP klasik seperti DVWA, sehingga hasil evaluasi akan lebih representatif terhadap kondisi sistem nyata.

Ketiga, disarankan untuk mengevaluasi pengaruh konfigurasi lanjutan (*advanced configuration*) terhadap performa WVS, seperti penggunaan *authenticated scanning*, pengaturan tingkat agresivitas pemindaian, penyesuaian *timeout*, serta integrasi *wordlist* tambahan. Penelitian ini menggunakan konfigurasi bawaan (*default*), sehingga eksplorasi konfigurasi lanjutan berpotensi meningkatkan nilai *recall* pada beberapa alat.

Keempat, untuk implementasi praktis di lingkungan organisasi, disarankan agar proses pemindaian tidak hanya mengandalkan satu alat saja. Hasil penelitian menunjukkan bahwa setiap WVS memiliki spesialisasi deteksi yang berbeda. Oleh karena itu, kombinasi beberapa alat, khususnya yang memiliki fokus berbeda (misalnya eksploitasi aktif dan analisis konfigurasi), dapat meningkatkan cakupan deteksi secara keseluruhan.

Dengan mempertimbangkan saran-saran tersebut, diharapkan penelitian selanjutnya dapat menghasilkan evaluasi yang lebih komprehensif dan memberikan kontribusi yang lebih luas terhadap pengembangan metode pengujian keamanan aplikasi web.